

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau(43) International Publication Date  
3 July 2003 (03.07.2003)

PCT

(10) International Publication Number  
WO 03/054219 A2

(51) International Patent Classification<sup>7</sup>: C12Q

(21) International Application Number: PCT/US02/41115

(22) International Filing Date:  
18 December 2002 (18.12.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/343,004	19 December 2001 (19.12.2001)	US
60/347,633	11 January 2002 (11.01.2002)	US
60/351,749	25 January 2002 (25.01.2002)	US
60/359,498	22 February 2002 (22.02.2002)	US

Yeun [KR/US]; 40 Westdale Avenue, Daly City, CA 94015 (US). LI, Joana, X. [US/US]; 1264 Geneva Avenue, San Francisco, CA 94112 (US). MARQUIS, Joseph, P. [US/US]; 4428 Lazy Lane, San Jose, CA 95135 (US). RAMKUMAR, Jayalaxmi [IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US). RICHARDSON, Thomas, W. [US/US]; 616 Canyon Road #107, Redwood City, CA 94062 (US). SPRAGUE, William, W. [US/US]; 611 13th Street # C, Sacramento, CA 95814 (US). SWARNAKAR, Anita [CA/US]; 8 Locksley Avenue #5D, San Francisco, CA 94122 (US). TANG, Y., Tom [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). CHAWLA, Narinder, K. [US/US]; 33 Union Square, #712, Union City, CA 94587 (US). WARREN, Bridget, A. [US/US]; 1810 S. El Camino Real #B103, Encinitas, CA 94024 (US). YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US).

(71) Applicant (for all designated States except US): INCYTE GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): AZIMZAI, Yalda [US/US]; 5518 Boulder Canyon Drive, Castro Valley, CA 94552 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). BECHA, Shanya, D. [US/US]; 21062 Gary Drive # 117, Castro Valley, CA 94546 (US). BOROWSKY, Mark, L. [US/US]; 122 Orchard Avenue, Redwood City, CA 94061 (US). BURFORD, Neil [GB/US]; 105 Wildwood Circle, Durham, CT 06422 (US). ELLIOTT, Vicki, S. [US/US]; 3770 Polton Place Way, San Jose, CA 95121 (US). EMERLING, Brooke, M. [US/US]; 1735 Woodland Avenue # 71, Palo Alto, CA 94303 (US). FORSYTHE, Ian, J. [US/US]; 308 Roble Avenue, Redwood City, CA 94061 (US). GORVAD, Ann, E. [US/US]; 369 Marie Common, Livermore, CA 94550 (US). GRIFFIN, Jennifer, A. [US/US]; 33691 Mello Way, Fremont, CA 94555 (US). KABLE, Amy, E. [US/US]; 2345 Polk Street #4, San Francisco, CA 94109 (US). KHARE, Reena [US/US]; 12650 Orella Court, Saratoga, CA 95070 (US). LAL, Preeti, G. [US/US]; P.O. Box 5142, Santa Clara, CA 95056 (US). LEE, Ernestine, A. [US/US]; 20523 Crow Creek Road, Castro Valley, CA 94552 (US). LEE, Soo,

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

## Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: NUCLEIC ACID-ASSOCIATED PROTEINS

(57) Abstract: Various embodiments of the invention provide human nucleic acid-associated proteins (NAAP) and polynucleotides which identify and encode NAAP. Embodiments of the invention also provide expression vectors, host cells, antibodies, agonists, and antagonists. Other embodiments provide methods for diagnosing, treating, or preventing disorders associated with aberrant expression of NAAP.



WO 03/054219 A2

## NUCLEIC ACID-ASSOCIATED PROTEINS

### TECHNICAL FIELD

The invention relates to novel nucleic acids, nucleic acid-associated proteins encoded by these  
5 nucleic acids, and to the use of these nucleic acids and proteins in the diagnosis, treatment, and  
prevention of cell proliferative, neurological, developmental, and autoimmune/inflammatory disorders,  
and infections. The invention also relates to the assessment of the effects of exogenous compounds  
on the expression of nucleic acids and nucleic acid-associated proteins.

### 10 BACKGROUND OF THE INVENTION

Multicellular organisms are comprised of diverse cell types that differ dramatically both in  
structure and function. The identity of a cell is determined by its characteristic pattern of gene  
expression, and different cell types express overlapping but distinctive sets of genes throughout  
development. Spatial and temporal regulation of gene expression is critical for the control of cell  
15 proliferation, cell differentiation, apoptosis, and other processes that contribute to organismal  
development. Furthermore, gene expression is regulated in response to extracellular signals that  
mediate cell-cell communication and coordinate the activities of different cell types. Appropriate gene  
regulation also ensures that cells function efficiently by expressing only those genes whose functions  
are required at a given time.

### 20 Transcription Factors

Transcriptional regulatory proteins are essential for the control of gene expression. Some of  
these proteins function as transcription factors that initiate, activate, repress, or terminate gene  
transcription. Transcription factors generally bind to the promoter, enhancer, and upstream regulatory  
regions of a gene in a sequence-specific manner, although some factors bind regulatory elements  
25 within or downstream of a gene coding region. Transcription factors may bind to a specific region of  
DNA singly or as a complex with other accessory factors. (Reviewed in Lewin, B. (1990) Genes IV,  
Oxford University Press, New York, NY, and Cell Press, Cambridge, MA, pp. 554-570.)

The double helix structure and repeated sequences of DNA create topological and chemical  
features which can be recognized by transcription factors. These features are hydrogen bond donor  
30 and acceptor groups, hydrophobic patches, major and minor grooves, and regular, repeated stretches  
of sequence which induce distinct bends in the helix. Typically, transcription factors recognize specific  
DNA sequence motifs of about 20 nucleotides in length. Multiple, adjacent transcription factor-binding  
motifs may be required for gene regulation.

Many transcription factors incorporate DNA-binding structural motifs which comprise either  $\alpha$  helices or  $\beta$  sheets that bind to the major groove of DNA. Four well-characterized structural motifs are helix-turn-helix, zinc finger, leucine zipper, and helix-loop-helix. Proteins containing these motifs may act alone as monomers, or they may form homo- or heterodimers that interact with DNA.

5       The helix-turn-helix motif consists of two  $\alpha$  helices connected at a fixed angle by a short chain of amino acids. One of the helices binds to the major groove. Helix-turn-helix motifs are exemplified by the homeobox motif which is present in homeodomain proteins. These proteins are critical for specifying the anterior-posterior body axis during development and are conserved throughout the animal kingdom. The Antennapedia and Ultrabithorax proteins of *Drosophila*  
10 *melanogaster* are prototypical homeodomain proteins. (Pabo, C.O. and R.T. Sauer (1992) Annu. Rev. Biochem. 61:1053-1095.)

      The zinc finger motif, which binds zinc ions, generally contains tandem repeats of about 30 amino acids consisting of periodically spaced cysteine and histidine residues. Examples of this sequence pattern, designated C2H2 and C3HC4 ("RING" finger), have been described. (Lewin,  
15 *supra*.) Zinc finger proteins each contain an  $\alpha$  helix and an antiparallel  $\beta$  sheet whose proximity and conformation are maintained by the zinc ion. Contact with DNA is made by the arginine preceding the  $\alpha$  helix and by the second, third, and sixth residues of the  $\alpha$  helix. Variants of the zinc finger motif include poorly defined cysteine-rich motifs which bind zinc or other metal ions. These motifs may not contain histidine residues and are generally nonrepetitive. The zinc finger motif may be repeated in a  
20 tandem array within a protein, such that the  $\alpha$  helix of each zinc finger in the protein makes contact with the major groove of the DNA double helix. This repeated contact between the protein and the DNA produces a strong and specific DNA-protein interaction. The strength and specificity of the interaction can be regulated by the number of zinc finger motifs within the protein. Though originally identified in DNA-binding proteins as regions that interact directly with DNA, zinc fingers occur in a  
25 variety of proteins that do not bind DNA (Lodish, H. et al. (1995) Molecular Cell Biology, Scientific American Books, New York, NY, pp. 447-451). For example, Galcheva-Gargova, Z. et al. (1996) Science 272:1797-1802) have identified zinc finger proteins that interact with various cytokine receptors.

      The C2H2-type zinc finger signature motif contains a 28 amino acid sequence, including 2  
30 conserved Cys and 2 conserved His residues in a C-2-C-12-H-3-H type motif. The motif generally occurs in multiple tandem repeats. A cysteine-rich domain including the motif Asp-His-His-Cys (DHHC-CRD) has been identified as a distinct subgroup of zinc finger proteins. The DHHC-CRD region has been implicated in growth and development. One DHHC-CRD mutant shows defective

function of Ras, a small membrane-associated GTP-binding protein that regulates cell growth and differentiation, while other DHHC-CRD proteins probably function in pathways not involving Ras (Bartels, D.J. et al. (1999) *Mol. Cell Biol.* 19:6775-6787).

Zinc-finger transcription factors are often accompanied by modular sequence motifs such as  
 5 the Kruppel-associated box (KRAB) and the SCAN domain. For example, the  
 hypoalphalipoproteinemia susceptibility gene ZNF202 encodes a SCAN box and a KRAB domain  
 followed by eight C2H2 zinc-finger motifs (Honer, C. et al. (2001) *Biochim. Biophys. Acta*  
 1517:441-448). The SCAN domain is a highly conserved, leucine-rich motif of approximately 60  
 amino acids found at the amino-terminal end of zinc finger transcription factors. SCAN domains are  
 10 most often linked to C2H2 zinc finger motifs through their carboxyl-terminal end. Biochemical binding  
 studies have established the SCAN domain as a selective hetero- and homotypic oligomerization  
 domain. SCAN domain-mediated protein complexes may function to modulate the biological function  
 of transcription factors (Schumacher, C. et al. (2000) *J. Biol. Chem.* 275:17173-17179).

The KRAB (Kruppel-associated box) domain is a conserved amino acid sequence spanning  
 15 approximately 75 amino acids and is found in almost one-third of the 300 to 700 genes encoding C2H2  
 zinc fingers. The KRAB domain is found N-terminally with respect to the finger repeats. The KRAB  
 domain is generally encoded by two exons; the KRAB-A region or box is encoded by one exon and  
 the KRAB-B region or box is encoded by a second exon. The function of the KRAB domain is the  
 repression of transcription. Transcription repression is accomplished by recruitment of either the  
 20 KRAB-associated protein-1, a transcriptional corepressor, or the KRAB-A interacting protein.  
 Proteins containing the KRAB domain are likely to play a regulatory role during development  
 (Williams, A.J. et al. (1999) *Mol. Cell Biol.* 19:8526-8535). A subgroup of highly related human  
 KRAB zinc finger proteins detectable in all human tissues is highly expressed in human T lymphoid  
 cells (Bellefroid, E.J. et al. (1993) *EMBO J.* 12:1363-1374). The ZNF85 KRAB zinc finger gene, a  
 25 member of the human ZNF91 family, is highly expressed in normal adult testis, in seminomas, and in  
 the NT2/D1 teratocarcinoma cell line (Poncelet, D.A. et al. (1998) *DNA Cell Biol.* 17:931-943).

The C4 motif is found in hormone-regulated proteins. The C4 motif generally includes only 2  
 repeats. A number of eukaryotic and viral proteins contain a conserved cysteine-rich domain of 40  
 to 60 residues (called C3HC4 zinc-finger or RING finger) that binds two atoms of zinc, and is  
 30 probably involved in mediating protein-protein interactions. The 3D "cross-brace" structure of the zinc  
 ligation system is unique to the RING domain. The spacing of the cysteines in such a domain is  
 C-x(2)-C-x(9 to 39)-C-x(1 to 3)-H-x(2 to 3)-C-x(2)-C-x(4 to 48)-C-x(2)-C. The PHD finger is a  
 C4HC3 zinc-finger-like motif found in nuclear proteins thought to be involved in chromatin-mediated



transcriptional regulation.

GATA-type transcription factors contain one or two zinc finger domains which bind specifically to a region of DNA that contains the consecutive nucleotide sequence GATA. NMR studies indicate that the zinc finger comprises two irregular anti-parallel  $\beta$  sheets and an  $\alpha$  helix, followed by a long loop to the C-terminal end of the finger (Ominchinski, J.G. (1993) Science 261:438-446). The helix and the loop connecting the two  $\beta$ -sheets contact the major groove of the DNA, while the C-terminal part, which determines the specificity of binding, wraps around into the minor groove.

The LIM motif consists of about 60 amino acid residues and contains seven conserved cysteine residues and a histidine within a consensus sequence (Schmeichel, K.L. and Beckerle, M.C. (1994) Cell 79:211-219). The LIM family includes transcription factors and cytoskeletal proteins which may be involved in development, differentiation, and cell growth. One example is actin-binding LIM protein, which may play roles in regulation of the cytoskeleton and cellular morphogenesis (Roof, D.J. et al. (1997) J. Cell Biol. 138:575-588). The N-terminal domain of actin-binding LIM protein has four double zinc finger motifs with the LIM consensus sequence. The C-terminal domain of actin-binding LIM protein shows sequence similarity to known actin-binding proteins such as dematin and villin. Actin-binding LIM protein binds to F-actin through its dematin-like C-terminal domain. The LIM domain may mediate protein-protein interactions with other LIM-binding proteins.

Myeloid cell development is controlled by tissue-specific transcription factors. Myeloid zinc finger proteins (MZF) include MZF-1 and MZF-2. MZF-1 functions in regulation of the development of neutrophilic granulocytes. A murine homolog MZF-2 is expressed in myeloid cells, particularly in the cells committed to the neutrophilic lineage. MZF-2 is down-regulated by G-CSF and appears to have a unique function in neutrophil development (Murai, K. et al. (1997) Genes Cells 2:581-591).

The leucine zipper motif comprises a stretch of amino acids rich in leucine which can form an amphipathic  $\alpha$  helix. This structure provides the basis for dimerization of two leucine zipper proteins. The region adjacent to the leucine zipper is usually basic, and upon protein dimerization, is optimally positioned for binding to the major groove. Proteins containing such motifs are generally referred to as bZIP transcription factors. The leucine zipper motif is found in the proto-oncogenes Fos and Jun, which comprise the heterodimeric transcription factor AP1 involved in cell growth and the determination of cell lineage (Papavassiliou, A.G. (1995) N. Engl. J. Med. 332:45-47).

The helix-loop-helix motif (HLH) consists of a short  $\alpha$  helix connected by a loop to a longer  $\alpha$  helix. The loop is flexible and allows the two helices to fold back against each other and to bind to DNA. The transcription factor Myc contains a prototypical HLH motif.

The NF-kappa-B/Rel signature defines a family of eukaryotic transcription factors involved in oncogenesis, embryonic development, differentiation and immune response. Most transcription factors containing the Rel homology domain (RHD) bind as dimers to a consensus DNA sequence motif termed kappa-B. Members of the Rel family share a highly conserved 300 amino acid domain termed the Rel homology domain. The characteristic Rel C-terminal domain is involved in gene activation and cytoplasmic anchoring functions. Proteins known to contain the RHD domain include vertebrate nuclear factor NF-kappa-B, which is a heterodimer of a DNA-binding subunit and the transcription factor p65, mammalian transcription factor RelB, and vertebrate proto-oncogene c-rel, a protein associated with differentiation and lymphopoiesis (Kabrun, N. and Enrietto, P.J. (1994) Semin. Cancer Biol. 5:103-112).

A DNA binding motif termed ARID (AT-rich interactive domain) distinguishes an evolutionarily conserved family of proteins. The approximately 100-residue ARID sequence is present in a series of proteins strongly implicated in the regulation of cell growth, development, and tissue-specific gene expression. ARID proteins include Bright (a regulator of B-cell-specific gene expression), dead ringer (involved in development), and MRF-2 (which represses expression from the cytomegalovirus enhancer) (Dallas, P.B. et al. (2000) Mol. Cell Biol. 20:3137-3146).

The ELM2 (Egl-27 and MTA1 homology 2) domain is found in metastasis-associated protein MTA1 and protein ER1. The *Caenorhabditis elegans* gene *egl-27* is required for embryonic patterning MTA1, a human gene with elevated expression in metastatic carcinomas, is a component of a protein complex with histone deacetylase and nucleosome remodelling activities (Solari, F. et al. (1999) Development 126:2483-2494). The ELM2 domain is usually found to the N terminus of a myb-like DNA binding domain. ELM2 is also found associated with an ARID DNA.

The Iroquois (Irx) family of genes are found in nematodes, insects and vertebrates. Irx genes usually occur in one or two genomic clusters of three genes each and encode transcriptional controllers that possess a characteristic homeodomain. The Irx genes function early in development to specify the identity of diverse territories of the body. Later in development in both *Drosophila* and vertebrates, the Irx genes function again to subdivide those territories into smaller domains. (For a review of Iroquois genes, see Cavodeassi, F. et al. (2001) Development 128:2847-2855.) For example, mouse and human Irx4 proteins are 83% conserved and their 63-aa homeodomain is more than 93% identical to that of the *Drosophila* Iroquois patterning genes. Irx4 transcripts are predominantly expressed in the cardiac ventricles. The homeobox gene Irx4 mediates ventricular differentiation during cardiac development (Bruneau, B.G. et al. (2000) Dev. Biol. 217:266-77).

Histidine triad (HIT) proteins share residues in distinctive dimeric, 10-stranded half-barrel

structures that form two identical purine nucleotide-binding sites. Hint (histidine triad nucleotide-binding protein)-related proteins, found in all forms of life, and fragile histidine triad (Fhit)-related proteins, found in animals and fungi, represent the two main branches of the HIT superfamily. Fhit homologs bind and cleave diadenosine polyphosphates. Fhit-Ap(n)A complexes  
5 appear to function in a proapoptotic tumor suppression pathway in epithelial tissues (Brenner C. et al. (1999) J. Cell Physiol. 181:179-187).

Most transcription factors contain characteristic DNA binding motifs, and variations on the above motifs and new motifs have been and are currently being characterized. (Faisst, S. and S. Meyer (1992) Nucleic Acids Res. 20:3-26.)

#### 10 Chromatin Associated Proteins

In the nucleus, DNA is packaged into chromatin, the compact organization of which limits the accessibility of DNA to transcription factors and plays a key role in gene regulation. (Lewin, *supra*, pp. 409-410.) The compact structure of chromatin is determined and influenced by chromatin-associated proteins such as the histones, the high mobility group (HMG) proteins, and the  
15 chromodomain proteins. There are five classes of histones, H1, H2A, H2B, H3, and H4, all of which are highly basic, low molecular weight proteins. The fundamental unit of chromatin, the nucleosome, consists of 200 base pairs of DNA associated with two copies each of H2A, H2B, H3, and H4. H1 links adjacent nucleosomes. HMG proteins are low molecular weight, non-histone proteins that may play a role in unwinding DNA and stabilizing single-stranded DNA. Chromodomain proteins play a  
20 key role in the formation of highly compacted heterochromatin, which is transcriptionally silent.

#### Diseases and Disorders Related to Gene Regulation

Many neoplastic disorders in humans can be attributed to inappropriate gene expression. Malignant cell growth may result from either excessive expression of tumor promoting genes or insufficient expression of tumor suppressor genes. (Cleary, M.L. (1992) Cancer Surv. 15:89-104.)

25 The zinc finger-type transcriptional regulator WT1 is a tumor-suppressor protein that is inactivated in children with Wilm's tumor. The oncogene bcl-6, which plays an important role in large-cell lymphoma, is also a zinc-finger protein (Papavassiliou, A.G. (1995) N. Engl. J. Med. 332:45-47). Chromosomal translocations may also produce chimeric loci that fuse the coding sequence of one gene with the regulatory regions of a second unrelated gene. Such an arrangement likely results in  
30 inappropriate gene transcription, potentially contributing to malignancy. In Burkitt's lymphoma, for example, the transcription factor Myc is translocated to the immunoglobulin heavy chain locus, greatly enhancing Myc expression and resulting in rapid cell growth leading to leukemia (Latchman, D.S. (1996) N. Engl. J. Med. 334:28-33).

In addition, the immune system responds to infection or trauma by activating a cascade of events that coordinate the progressive selection, amplification, and mobilization of cellular defense mechanisms. A complex and balanced program of gene activation and repression is involved in this process. However, hyperactivity of the immune system as a result of improper or insufficient  
 5 regulation of gene expression may result in considerable tissue or organ damage. This damage is well-documented in immunological responses associated with arthritis, allergens, heart attack, stroke, and infections. (Isselbacher et al. Harrison's Principles of Internal Medicine, 13/e, McGraw Hill, Inc. and Teton Data Systems Software, 1996.) The causative gene for autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED) was recently isolated and found to encode a protein  
 10 with two PHD-type zinc finger motifs (Bjorses, P. et al. (1998) *Hum. Mol. Genet.* 7:1547-1553).

Furthermore, the generation of multicellular organisms is based upon the induction and coordination of cell differentiation at the appropriate stages of development. Central to this process is differential gene expression, which confers the distinct identities of cells and tissues throughout the body. Failure to regulate gene expression during development could result in developmental disorders.  
 15 Human developmental disorders caused by mutations in zinc finger-type transcriptional regulators include: urogenital developmental abnormalities associated with WT1; Greig cephalopolysyndactyly, Pallister-Hall syndrome, and postaxial polydactyly type A (GLI3), and Townes-Brocks syndrome, characterized by anal, renal, limb, and ear abnormalities (SALL1) (Engelkamp, D. and V. van : Heyningen (1996) *Curr. Opin. Genet. Dev.* 6:334-342; Kohlhase, J. et al. (1999) *Am. J. Hum. Genet.*  
 20 64:435-445).

Human acute leukemias involve reciprocal chromosome translocations that fuse the ALL-1 gene located at chromosome region 11q23 to a series of partner genes positioned on a variety of human chromosomes. The fused genes encode chimeric proteins. The AF17 gene encodes a protein of 1093 amino acids, containing a leucine-zipper dimerization motif located 3' of the fusion point and a  
 25 cysteine-rich domain at the N terminus that shows homology to a domain within the protein Br140 (peregrin) (Prasad R. et al. (1994) *Proc. Natl. Acad. Sci. U S A* 91:8107-8111).

## SYNTHESIS OF NUCLEIC ACIDS

### Polymerases

DNA and RNA replication are critical processes for cell replication and function. DNA and  
 30 RNA replication are mediated by the enzymes DNA and RNA polymerase, respectively, by a "templating" process in which the nucleotide sequence of a DNA or RNA strand is copied by complementary base-pairing into a complementary nucleic acid sequence of either DNA or RNA. However, there are fundamental differences between the two processes.

DNA polymerase catalyzes the stepwise addition of a deoxyribonucleotide to the 3'-OH end of a polynucleotide strand (the primer strand) that is paired to a second (template) strand. The new DNA strand therefore grows in the 5' to 3' direction (Alberts, B. et al. (1994) The Molecular Biology of the Cell, Garland Publishing Inc., New York, NY, pp 251-254). The substrates for the

5 polymerization reaction are the corresponding deoxynucleotide triphosphates which must base-pair with the correct nucleotide on the template strand in order to be recognized by the polymerase. Because DNA exists as a double-stranded helix, each of the two strands may serve as a template for the formation of a new complementary strand. Each of the two daughter cells of a dividing cell therefore inherits a new DNA double helix containing one old and one new strand. Thus, DNA is said

10 to be replicated "semiconservatively" by DNA polymerase. In addition to the synthesis of new DNA, DNA polymerase is also involved in the repair of damaged DNA as discussed below under "Ligases."

In contrast to DNA polymerase, RNA polymerase uses a DNA template strand to "transcribe" DNA into RNA using ribonucleotide triphosphates as substrates. Like DNA polymerization, RNA polymerization proceeds in a 5' to 3' direction by addition of a ribonucleoside

15 monophosphate to the 3'-OH end of a growing RNA chain. DNA transcription generates messenger RNAs (mRNA) that carry information for protein synthesis, as well as the transfer, ribosomal, and other RNAs that have structural or catalytic functions. In eukaryotes, three discrete RNA polymerases synthesize the three different types of RNA (Alberts, *supra*, pp. 367-368). RNA polymerase I makes the large ribosomal RNAs, RNA polymerase II makes the mRNAs that will be

20 translated into proteins, and RNA polymerase III makes a variety of small, stable RNAs, including 5S ribosomal RNA and the transfer RNAs (tRNA). In all cases, RNA synthesis is initiated by binding of the RNA polymerase to a promoter region on the DNA and synthesis begins at a start site within the promoter. Synthesis is completed at a stop (termination) signal in the DNA whereupon both the polymerase and the completed RNA chain are released.

## 25 Ligases

DNA repair is the process by which accidental base changes, such as those produced by oxidative damage, hydrolytic attack, or uncontrolled methylation of DNA, are corrected before replication or transcription of the DNA can occur. Because of the efficiency of the DNA repair process, fewer than one in a thousand accidental base changes causes a mutation (Alberts, *supra*, pp.

30 245-249). The three steps common to most types of DNA repair are (1) excision of the damaged or altered base or nucleotide by DNA nucleases, (2) insertion of the correct nucleotide in the gap left by the excised nucleotide by DNA polymerase using the complementary strand as the template and, (3) sealing the break left between the inserted nucleotide(s) and the existing DNA strand by DNA ligase.

In the last reaction, DNA ligase uses the energy from ATP hydrolysis to activate the 5' end of the broken phosphodiester bond before forming the new bond with the 3'-OH of the DNA strand. In Bloom's syndrome, an inherited human disease, individuals are partially deficient in DNA ligation and consequently have an increased incidence of cancer (Alberts, *supra* p. 247).

## 5 Nucleases

Nucleases comprise enzymes that hydrolyze both DNA (DNase) and RNA (RNase). They serve different purposes in nucleic acid metabolism. Nucleases hydrolyze the phosphodiester bonds between adjacent nucleotides either at internal positions (endonucleases) or at the terminal 3' or 5' nucleotide positions (exonucleases). A DNA exonuclease activity in DNA polymerase, for example, 10 serves to remove improperly paired nucleotides attached to the 3'-OH end of the growing DNA strand by the polymerase and thereby serves a "proofreading" function. As mentioned above, DNA endonuclease activity is involved in the excision step of the DNA repair process.

RNases also serve a variety of functions. For example, RNase P is a ribonucleoprotein enzyme which cleaves the 5' end of pre-tRNAs as part of their maturation process. RNase H digests 15 the RNA strand of an RNA/DNA hybrid. Such hybrids occur in cells invaded by retroviruses, and RNase H is an important enzyme in the retroviral replication cycle. Pancreatic RNase secreted by the pancreas into the intestine hydrolyzes RNA present in ingested foods. RNase activity in serum and cell extracts is elevated in a variety of cancers and infectious diseases (Schein, C.H. (1997) *Nat. Biotechnol.* 15:529-536). Regulation of RNase activity is being investigated as a means to control 20 tumor angiogenesis, allergic reactions, viral infection and replication, and fungal infections.

## MODIFICATION OF NUCLEIC ACIDS

### Methylases

Methylation of specific nucleotides occurs in both DNA and RNA, and serves different functions in the two macromolecules. Methylation of cytosine residues to form 5-methyl cytosine in 25 DNA occurs specifically in CG sequences which are base-paired with one another in the DNA double-helix. The pattern of methylation is passed from generation to generation during DNA replication by an enzyme called "maintenance methylase" that acts preferentially on those CG sequences that are base-paired with a CG sequence that is already methylated. Such methylation appears to distinguish active from inactive genes by preventing the binding of regulatory proteins that 30 "turn on" the gene, but permitting the binding of proteins that inactivate the gene (Alberts, *supra* pp. 448-451). In RNA metabolism, "tRNA methylase" produces one of several nucleotide modifications in tRNA that affect the conformation and base-pairing of the molecule and facilitate the recognition of the appropriate mRNA codons by specific tRNAs. The primary methylation pattern is the

dimethylation of guanine residues to form N,N-dimethyl guanine.

### **Helicases and Single-stranded Binding Proteins**

Helicases are enzymes that destabilize and unwind double helix structures in both DNA and RNA. Since DNA replication occurs more or less simultaneously on both strands, the two strands  
5 must first separate to generate a replication "fork" for DNA polymerase to act on. Two types of replication proteins contribute to this process, DNA helicases and single-stranded binding proteins. DNA helicases hydrolyze ATP and use the energy of hydrolysis to separate the DNA strands. Single-stranded binding proteins (SSBs) then bind to the exposed DNA strands, without covering the bases, thereby temporarily stabilizing them for templating by the DNA polymerase (Alberts, *supra*, pp.  
10 255-256).

RNA helicases also alter and regulate RNA conformation and secondary structure. Like the DNA helicases, RNA helicases utilize energy derived from ATP hydrolysis to destabilize and unwind RNA duplexes. The most well-characterized and ubiquitous family of RNA helicases is the DEAD-box family, so named for the conserved B-type ATP-binding motif which is diagnostic of proteins in  
15 this family. Over 40 DEAD-box helicases have been identified in organisms as diverse as bacteria, insects, yeast, amphibians, mammals, and plants. DEAD-box helicases function in diverse processes such as translation initiation, splicing, ribosome assembly, and RNA editing, transport, and stability. Examples of these RNA helicases include yeast Drs1 protein, which is involved in ribosomal RNA processing; yeast TIF1 and TIF2 and mammalian eIF-4A, which are essential to the initiation of RNA  
20 translation; and human p68 antigen, which regulates cell growth and division (Ripmaster, T.L. et al. (1992) Proc. Natl. Acad. Sci. USA 89:11131-11135; Chang, T.-H. et al. (1990) Proc. Natl. Acad. Sci. USA 87:1571-1575). These RNA helicases demonstrate strong sequence homology over a stretch of some 420 amino acids. Included among these conserved sequences are the consensus sequence for the A motif of an ATP binding protein; the "DEAD box" sequence, associated with ATPase activity;  
25 the sequence SAT, associated with the actual helicase unwinding region; and an octapeptide consensus sequence, required for RNA binding and ATP hydrolysis (Pause, A. et al. (1993) Mol. Cell Biol. 13:6789-6798). Differences outside of these conserved regions are believed to reflect differences in the functional roles of individual proteins (Chang, T.H. et al. (1990) Proc. Natl. Acad. Sci. USA 87:1571-1575).

30 Some DEAD-box helicases play tissue- and stage-specific roles in spermatogenesis and embryogenesis. Overexpression of the DEAD-box 1 protein (DDX1) may play a role in the progression of neuroblastoma (Nb) and retinoblastoma (Rb) tumors (Godbout, R. et al. (1998) J. Biol. Chem. 273:21161-21168). These observations suggest that DDX1 may promote or enhance tumor

progression by altering the normal secondary structure and expression levels of RNA in cancer cells. Other DEAD-box helicases have been implicated either directly or indirectly in tumorigenesis. (Discussed in Godbout, *supra*.) For example, murine p68 is mutated in ultraviolet light-induced tumors, and human DDX6 is located at a chromosomal breakpoint associated with B-cell lymphoma.

- 5 Similarly, a chimeric protein comprised of DDX10 and NUP98, a nucleoporin protein, may be involved in the pathogenesis of certain myeloid malignancies.

### Topoisomerases

- Besides the need to separate DNA strands prior to replication, the two strands must be “unwound” from one another prior to their separation by DNA helicases. This function is performed
- 10 by proteins known as DNA topoisomerases. DNA topoisomerase effectively acts as a reversible nuclease that hydrolyzes a phosphodiesterase bond in a DNA strand, permits the two strands to rotate freely about one another to remove the strain of the helix, and then rejoins the original phosphodiester bond between the two strands. Topoisomerases are essential enzymes responsible for the topological rearrangement of DNA brought about by transcription, replication, chromatin formation,
- 15 recombination, and chromosome segregation. Superhelical coils are introduced into DNA by the passage of processive enzymes such as RNA polymerase, or by the separation of DNA strands by a helicase prior to replication. Knotting and concatenation can occur in the process of DNA synthesis, storage, and repair. All topoisomerases work by breaking a phosphodiester bond in the ribose-phosphate backbone of DNA. A catalytic tyrosine residue on the enzyme makes a nucleophilic attack
- 20 on the scissile phosphodiester bond, resulting in a reaction intermediate in which a covalent bond is formed between the enzyme and one end of the broken strand. A tyrosine-DNA phosphodiesterase functions in DNA repair by hydrolyzing this bond in occasional dead-end topoisomerase I-DNA intermediates (Pouliot, J.J. et al. (1999) Science 286:552-555).

- Two types of DNA topoisomerase exist, types I and II. Type I topoisomerases work as
- 25 monomers, making a break in a single strand of DNA while type II topoisomerases, working as homodimers, cleave both strands. DNA Topoisomerase I causes a single-strand break in a DNA helix to allow the rotation of the two strands of the helix about the remaining phosphodiester bond in the opposite strand. DNA topoisomerase II causes a transient break in both strands of a DNA helix where two double helices cross over one another. This type of topoisomerase can efficiently separate
- 30 two interlocked DNA circles (Alberts, *supra*, pp.260-262). Type II topoisomerases are largely confined to proliferating cells in eukaryotes, such as cancer cells. For this reason they are targets for anticancer drugs. Topoisomerase II has been implicated in multi-drug resistance (MDR) as it appears to aid in the repair of DNA damage inflicted by DNA binding agents such as doxorubicin and



vincristine.

The topoisomerase I family includes topoisomerases I and III (topo I and topo III). The crystal structure of human topoisomerase I suggests that rotation about the intact DNA strand is partially controlled by the enzyme. In this "controlled rotation" model, protein-DNA interactions limit the rotation, which is driven by torsional strain in the DNA (Stewart, L. et al. (1998) *Science* 281:1534-1541). Structurally, topo I can be recognized by its catalytic tyrosine residue and a number of other conserved residues in the active site region. Topo I is thought to function during transcription. Two topo IIIs are known in humans, and they are homologous to prokaryotic topoisomerase I, with a conserved tyrosine and active site signature specific to this family. Topo III has been suggested to play a role in meiotic recombination. A mouse topo III is highly expressed in testis tissue and its expression increases with the increase in the number of cells in pachytene (Seki, T. et al. (1998) *J. Biol. Chem.* 273:28553-28556).

The topoisomerase II family includes two isozymes (II $\alpha$  and II $\beta$ ) encoded by different genes. Topo II cleaves double stranded DNA in a reproducible, nonrandom fashion, preferentially in an AT rich region, but the basis of cleavage site selectivity is not known. Structurally, topo II is made up of four domains, the first two of which are structurally similar and probably distantly homologous to similar domains in eukaryotic topo I. The second domain bears the catalytic tyrosine, as well as a highly conserved pentapeptide. The II $\alpha$  isoform appears to be responsible for unlinking DNA during chromosome segregation. Cell lines expressing II $\alpha$  but not II $\beta$  suggest that II $\beta$  is dispensable in cellular processes; however, II $\beta$  knockout mice died perinatally due to a failure in neural development. That the major abnormalities occurred in predominantly late developmental events (neurogenesis) suggests that II $\beta$  is needed not at mitosis, but rather during DNA repair (Yang, X. et al. (2000) *Science* 287:131-134).

Topoisomerases have been implicated in a number of disease states, and topoisomerase poisons have proven to be effective anti-tumor drugs for some human malignancies. Topo I is mislocalized in Fanconi's anemia, and may be involved in the chromosomal breakage seen in this disorder (Wunder, E. (1984) *Hum. Genet.* 68:276-281). Overexpression of a truncated topo III in ataxia-telangiectasia (A-T) cells partially suppresses the A-T phenotype, probably through a dominant negative mechanism. This suggests that topo III is deregulated in A-T (Fritz, E. et al. (1997) *Proc. Natl. Acad. Sci. USA* 94:4538-4542). Topo III also interacts with the Bloom's Syndrome gene product, and has been suggested to have a role as a tumor suppressor (Wu, L. et al. (2000) *J. Biol. Chem.* 275:9636-9644). Aberrant topo II activity is often associated with cancer or increased cancer risk. Greatly lowered topo II activity has been found in some, but not all A-T cell lines (Mohamed, R.

et al. (1987) Biochem. Biophys. Res. Commun. 149:233-238). On the other hand, topo II can break DNA in the region of the A-T gene (ATM), which controls all DNA damage-responsive cell cycle checkpoints (Kaufmann, W.K. (1998) Proc. Soc. Exp. Biol. Med. 217:327-334). The ability of topoisomerases to break DNA has been used as the basis of antitumor drugs. Topoisomerase poisons  
5 act by increasing the number of dead-end covalent DNA-enzyme complexes in the cell, ultimately triggering cell death pathways (Fortune, J.M. and N. Osheroff (2000) Prog. Nucleic Acid Res. Mol. Biol. 64:221-253; Guichard, S.M. and M.K. Danks (1999) Curr. Opin. Oncol. 11:482-489). Antibodies against topo I are found in the serum of systemic sclerosis patients, and the levels of the antibody may be used as a marker of pulmonary involvement in the disease (Diot, E. et al. (1999) Chest 116:715-  
10 720). Finally, the DNA binding region of human topo I has been used as a DNA delivery vehicle for gene therapy (Chen, T.Y. et al. (2000) Appl. Microbiol. Biotechnol. 53:558-567).

### Recombinases

Genetic recombination is the process of rearranging DNA sequences within an organism's genome to provide genetic variation for the organism in response to changes in the environment.  
15 DNA recombination allows variation in the particular combination of genes present in an individual's genome, as well as the timing and level of expression of these genes. (See Alberts, *supra* pp. 263-273.) Two broad classes of genetic recombination are commonly recognized, general recombination and site-specific recombination. General recombination involves genetic exchange between any homologous pair of DNA sequences usually located on two copies of the same chromosome. The  
20 process is aided by enzymes, recombinases, that "nick" one strand of a DNA duplex more or less randomly and permit exchange with a complementary strand on another duplex. The process does not normally change the arrangement of genes in a chromosome. In site-specific recombination, the recombinase recognizes specific nucleotide sequences present in one or both of the recombining molecules. Base-pairing is not involved in this form of recombination and therefore it does not require  
25 DNA homology between the recombining molecules. Unlike general recombination, this form of recombination can alter the relative positions of nucleotide sequences in chromosomes.

### **RNA METABOLISM**

Ribonucleic acid (RNA) is a linear single-stranded polymer of four nucleotides, ATP, CTP, UTP, and GTP. In most organisms, RNA is transcribed as a copy of deoxyribonucleic acid (DNA),  
30 the genetic material of the organism. In retroviruses RNA rather than DNA serves as the genetic material. RNA copies of the genetic material encode proteins or serve various structural, catalytic, or regulatory roles in organisms. RNA is classified according to its cellular localization and function. Messenger RNAs (mRNAs) encode polypeptides. Ribosomal RNAs (rRNAs) are assembled, along

with ribosomal proteins, into ribosomes, which are cytoplasmic particles that translate mRNA into polypeptides. Transfer RNAs (tRNAs) are cytosolic adaptor molecules that function in mRNA translation by recognizing both an mRNA codon and the amino acid that matches that codon.

Heterogeneous nuclear RNAs (hnRNAs) include mRNA precursors and other nuclear RNAs of

5 various sizes. Small nuclear RNAs (snRNAs) are a part of the nuclear spliceosome complex that removes intervening, non-coding sequences (introns) and rejoins exons in pre-mRNAs.

Proteins are associated with RNA during its transcription from DNA, RNA processing, and translation of mRNA into protein. Proteins are also associated with RNA as it is used for structural, catalytic, and regulatory purposes.

## 10 **RNA Processing**

Ribosomal RNAs (rRNAs) are assembled, along with ribosomal proteins, into ribosomes, which are cytoplasmic particles that translate messenger RNA (mRNA) into polypeptides. The eukaryotic ribosome is composed of a 60S (large) subunit and a 40S (small) subunit, which together form the 80S ribosome. In addition to the 18S, 28S, 5S, and 5.8S rRNAs, ribosomes contain from 50  
15 to over 80 different ribosomal proteins, depending on the organism. Ribosomal proteins are classified according to which subunit they belong (i.e., L, if associated with the large 60S large subunit or S if associated with the small 40S subunit). *E. coli* ribosomes have been the most thoroughly studied and contain 50 proteins, many of which are conserved in all life forms. The structures of nine ribosomal proteins have been solved to less than 3.0D resolution (i.e., S5, S6, S17, L1, L6, L9, L12, L14, L30),  
20 revealing common motifs, such as b-a-b protein folds in addition to acidic and basic RNA-binding motifs positioned between b-strands. Most ribosomal proteins are believed to contact rRNA directly (reviewed in Liljas, A. and Garber, M. (1995) Curr. Opin. Struct. Biol. 5:721-727; see also Woodson, S.A. and Leontis, N.B. (1998) Curr. Opin. Struct. Biol. 8:294-300; Ramakrishnan, V. and White, S.W. (1998) Trends Biochem. Sci. 23:208-212).

25 Ribosomal proteins may undergo post-translational modifications or interact with other ribosome-associated proteins to regulate translation. For example, the highly homologous 40S ribosomal protein S6 kinases (S6K1 and S6K2) play a key role in the regulation of cell growth by controlling the biosynthesis of translational components which make up the protein synthetic apparatus (including the ribosomal proteins). In the case of S6K1, at least eight phosphorylation sites are  
30 believed to mediate kinase activation in a hierarchical fashion (Dufner and Thomas (1999) Exp. Cell. Res. 253:100-109). Some of the ribosomal proteins, including L1, also function as translational repressors by binding to polycistronic mRNAs encoding ribosomal proteins (reviewed in Liljas, *supra* and Garber, *supra*).

Recent evidence suggests that a number of ribosomal proteins have secondary functions independent of their involvement in protein biosynthesis. These proteins function as regulators of cell proliferation and, in some instances, as inducers of cell death. For example, the expression of human ribosomal protein L13a has been shown to induce apoptosis by arresting cell growth in the G2/M  
5 phase of the cell cycle. Inhibition of expression of L13a induces apoptosis in target cells, which suggests that this protein is necessary, in the appropriate amount, for cell survival. Similar results have been obtained in yeast where inactivation of yeast homologues of L13a, rp22 and rp23, results in severe growth retardation and death. A closely related ribosomal protein, L7, arrests cells in G1 and also induces apoptosis. Thus, it appears that a subset of ribosomal proteins may function as cell cycle  
10 checkpoints and compose a new family of cell proliferation regulators.

Mapping of individual ribosomal proteins on the surface of intact ribosomes is accomplished using 3D immunocryoelectronmicroscopy, whereby antibodies raised against specific ribosomal proteins are visualized. Progress has been made toward the mapping of L1, L7, and L12 while the structure of the intact ribosome has been solved to only 20-25D resolution and inconsistencies exist  
15 among different crude structures (Frank, J. (1997) *Curr. Opin. Struct. Biol.* 7:266-272).

Three distinct sites have been identified on the ribosome. The aminoacyl-tRNA acceptor site (A site) receives charged tRNAs (with the exception of the initiator-tRNA). The peptidyl-tRNA site (P site) binds the nascent polypeptide as the amino acid from the A site is added to the elongating chain. Deacylated tRNAs bind in the exit site (E site) prior to their release from the ribosome. The  
20 structure of the ribosome is reviewed in Stryer, L. (1995) Biochemistry, W.H. Freeman and Company, New York NY, pp. 888-9081; Lodish, H. et al. (1995) Molecular Cell Biology, Scientific American Books, New York NY, pp. 119-138; and Lewin, B (1997) Genes VI, Oxford University Press, Inc. New York, NY).

Various proteins are necessary for processing of transcribed RNAs in the nucleus. Pre-  
25 mRNA processing steps include capping at the 5' end with methylguanosine, polyadenylating the 3' end, and splicing to remove introns. The primary RNA transcript from DNA is a faithful copy of the gene containing both exon and intron sequences, and the latter sequences must be cut out of the RNA transcript to produce a mRNA that codes for a protein. This "splicing" of the mRNA sequence takes place in the nucleus with the aid of a large, multicomponent ribonucleoprotein complex known as a  
30 spliceosome. The spliceosomal complex is comprised of five small nuclear ribonucleoprotein particles (snRNPs) designated U1, U2, U4, U5, and U6. Each snRNP contains a single species of snRNA and about ten proteins. The RNA components of some snRNPs recognize and base-pair with intron consensus sequences. The protein components mediate spliceosome assembly and the splicing

reaction. Autoantibodies to snRNP proteins are found in the blood of patients with systemic lupus erythematosus (Stryer, L. (1995) Biochemistry, W.H. Freeman and Company, New York NY, p. 863).

Several splicing regulatory proteins have been identified in Drosophila. Human (HsSWAP) and mouse (MmSWAP) homologs of the suppressor-of-white-apricot (su(wa)) gene have been cloned and characterized. HsSWAP and MmSWAP both have five highly homologous regions to su(wa), including an arginine/serine-rich domain and two repeated modules that are homologous to regions in the constitutive splicing factor, SPP91/PRP21. Mammalian SWAP mRNAs are alternatively spliced at the same splice sites as in Drosophila. The splice junctions of the Drosophila and mammalian regulated introns are conserved. Thus, research suggests that the mammalian SWAP gene functions as a vertebrate alternative splicing regulator (Denhez, F. and Lafyatis, R. (1994) *Biol. Chem.* 269:16170-16179).

Serine- and arginine-rich pre-mRNA splicing factors (SR proteins) are phosphorylated before they regulate splicing events. SRp86 (SR-related protein of 86 kDa) is a novel SR protein containing a single amino-terminal RNA recognition motif and two carboxy-terminal domains rich in serine-arginine (SR) dipeptides. SRp86 activates splicing in the presence of SRp20. However, it inhibits the in vitro and in vivo activation of specific splice sites by SR proteins, including ASF/SF2, SC35, and SRp55. Research suggests that pairwise combination of SRp86 with specific SR proteins leads to altered splicing efficiency and differential splice site selection (Barnard, D.C. and Patton, J.G. (2000) *Mol. Cell. Biol.* 20:3049-3057).

Heterogeneous nuclear ribonucleoproteins (hnRNPs) have been identified that have roles in splicing, exporting of the mature RNAs to the cytoplasm, and mRNA translation (Biamonti, G. et al. (1998) *Clin. Exp. Rheumatol.* 16:317-326). Some examples of hnRNPs include the yeast proteins Hrp1p, involved in cleavage and polyadenylation at the 3' end of the RNA; Cbp80p, involved in capping the 5' end of the RNA; and Npl3p, a homolog of mammalian hnRNP A1, involved in export of mRNA from the nucleus (Shen, E.C. et al. (1998) *Genes Dev.* 12:679-691). HnRNPs have been shown to be important targets of the autoimmune response in rheumatic diseases (Biamonti, *supra*).

Many snRNP and hnRNP proteins are characterized by an RNA recognition motif (RRM). (Reviewed in Birney, E. et al. (1993) *Nucleic Acids Res.* 21:5803-5816.) The RRM is about 80 amino acids in length and forms four  $\beta$ -strands and two  $\alpha$ -helices arranged in an  $\alpha/\beta$  sandwich. The RRM contains a core RNP-1 octapeptide motif along with surrounding conserved sequences. In addition to snRNP proteins, examples of RNA-binding proteins which contain the above motifs include heteronuclear ribonucleoproteins which stabilize nascent RNA and factors which regulate alternative

splicing. Alternative splicing factors include developmentally regulated proteins, specific examples of which have been identified in lower eukaryotes such as *Drosophila melanogaster* and *Caenorhabditis elegans*. These proteins play key roles in developmental processes such as pattern formation and sex determination, respectively. (See, for example, Hodgkin, J. et al. (1994)

5 Development 120:3681-3689.)

The 3' ends of most eukaryote mRNAs are also posttranscriptionally modified by polyadenylation. Polyadenylation proceeds through two enzymatically distinct steps: (i) the endonucleolytic cleavage of nascent mRNAs at *cis*-acting polyadenylation signals in the 3'-untranslated (non-coding) region and (ii) the addition of a poly(A) tract to the 5' mRNA fragment.

10 The presence of *cis*-acting RNA sequences is necessary for both steps. These sequences include 5'-AAUAAA-3' located 10-30 nucleotides upstream of the cleavage site and a less well-conserved GU- or U-rich sequence element located 10-30 nucleotides downstream of the cleavage site. Cleavage stimulation factor (CstF), cleavage factor I (CF I), and cleavage factor II (CF II) are involved in the cleavage reaction while cleavage and polyadenylation specificity factor (CPSF) and poly(A)  
15 polymerase (PAP) are necessary for both cleavage and polyadenylation. An additional enzyme, poly(A)-binding protein II (PAB II), promotes poly(A) tract elongation (Rüegsegger, U. et al. (1996) J. Biol. Chem. 271:6107-6113; and references within).

## TRANSLATION

Correct translation of the genetic code depends upon each amino acid forming a linkage with  
20 the appropriate transfer RNA (tRNA). The aminoacyl-tRNA synthetases (aaRSs) are essential proteins found in all living organisms. The aaRSs are responsible for the activation and correct attachment of an amino acid with its cognate tRNA, as the first step in protein biosynthesis. Prokaryotic organisms have at least twenty different types of aaRSs, one for each different amino acid, while eukaryotes usually have two aaRSs, a cytosolic form and a mitochondrial form, for each  
25 different amino acid. The 20 aaRS enzymes can be divided into two structural classes. Class I enzymes add amino acids to the 2' hydroxyl at the 3' end of tRNAs while Class II enzymes add amino acids to the 3' hydroxyl at the 3' end of tRNAs. Each class is characterized by a distinctive topology of the catalytic domain. Class I enzymes contain a catalytic domain based on the nucleotide-binding Rossman 'fold'. In particular, a consensus tetrapeptide motif is highly conserved (Prosite Document  
30 PDOC00161, Aminoacyl-transfer RNA synthetases class-I signature). Class I enzymes are specific for arginine, cysteine, glutamic acid, glutamine, isoleucine, leucine, methionine, tyrosine, tryptophan, and valine. Class II enzymes contain a central catalytic domain, which consists of a seven-stranded antiparallel  $\beta$ -sheet domain, as well as N- and C- terminal regulatory domains. Class II enzymes are

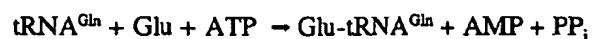
separated into two groups based on the heterodimeric or homodimeric structure of the enzyme; the latter group is further subdivided by the structure of the N- and C-terminal regulatory domains (Hartlein, M. and Cusack, S. (1995) *J. Mol. Evol.* 40:519-530). Class II enzymes are specific for alanine, asparagine, aspartic acid, glycine, histidine, lysine, phenylalanine, proline, serine, and threonine.

5 Certain aaRSs also have editing functions. IleRS, for example, can misactivate valine to form Val-tRNA<sup>Ile</sup>, but this product is cleared by a hydrolytic activity that destroys the mischarged product. This editing activity is located within a second catalytic site found in the connective polypeptide 1 region (CP1), a long insertion sequence within the Rossman fold domain of Class I enzymes (Schimmel, P. et al. (1998) *FASEB J.* 12:1599-1609). AaRSs also play a role in tRNA processing. It  
10 has been shown that mature tRNAs are charged with their respective amino acids in the nucleus before export to the cytoplasm, and charging may serve as a quality control mechanism to insure the tRNAs are functional (Martinis, S.A. et al. (1999) *EMBO J.* 18:4591-4596).

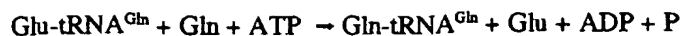
Under optimal conditions, polypeptide synthesis proceeds at a rate of approximately 40 amino acid residues per second. The rate of misincorporation during translation is on the order of  $10^{-4}$  and is  
15 primarily the result of aminoacyl-t-RNAs being charged with the incorrect amino acid. Incorrectly charged tRNA are toxic to cells as they result in the incorporation of incorrect amino acid residues into an elongating polypeptide. The rate of translation is presumed to be a compromise between the optimal rate of elongation and the need for translational fidelity. Mathematical calculations predict that  $10^{-4}$  is indeed the maximum acceptable error rate for protein synthesis in a biological system (reviewed  
20 in Stryer, *supra*; and Watson, J. et al. (1987) The Benjamin/Cummings Publishing Co., Inc. Menlo Park, CA). A particularly error prone aminoacyl-tRNA charging event is the charging of tRNA<sup>Gln</sup> with Gln. A mechanism exists for the correction of this mischarging event which likely has its origins in evolution. Gln was among the last of the 20 naturally occurring amino acids used in polypeptide synthesis to appear in nature. Gram positive eubacteria, cyanobacteria, Archeae, and eukaryotic  
25 organelles possess a noncanonical pathway for the synthesis of Gln-tRNA<sup>Gln</sup> based on the transformation of Glu-tRNA<sup>Gln</sup> (synthesized by Glu-tRNA synthetase, GluRS) using the enzyme Glu-tRNA<sup>Gln</sup> amidotransferase (Glu-AdT). The reactions involved in the transamidation pathway are as follows (Curnow, A.W. et al. (1997) *Nucleic Acids Symposium* 36:2-4):

30

GluRS



Glu-AdT



A similar enzyme, Asp-tRNA<sup>Asn</sup> amidotransferase, exists in Archaea, which transforms Asp-tRNA<sup>Asn</sup> to Asn-tRNA<sup>Asn</sup>. Formylase, the enzyme that transforms Met-tRNA<sup>Met</sup> to fMet-tRNA<sup>Met</sup> in eubacteria, is likely to be a related enzyme. A hydrolytic activity has also been identified that destroys  
 5 mischarged Val-tRNA<sup>Ile</sup> (Schimmel, P. et al. (1998) FASEB J. 12:1599-1609). One likely scenario for the evolution of Glu-AdT in primitive life forms is the absence of a specific glutamyl-tRNA synthetase (GlnRS), requiring an alternative pathway for the synthesis of Gln-tRNA<sup>Gln</sup>. In fact, deletion of the Glu-AdT operon in Gram positive bacteria is lethal (Curnow, A.W. et al. (1997) Proc. Natl. Acad. Sci. USA 94:11819-11826). The existence of GluRS activity in other organisms has been  
 10 inferred by the high degree of conservation in translation machinery in nature; however, GluRS has not been identified in all organisms, including *Homo sapiens*. Such an enzyme would be responsible for ensuring translational fidelity and reducing the synthesis of defective polypeptides.

In addition to their function in protein synthesis, specific aminoacyl tRNA synthetases also  
 1 play roles in cellular fidelity, RNA splicing, RNA trafficking, apoptosis, and transcriptional and  
 15 translational regulation. For example, human tyrosyl-tRNA synthetase can be proteolytically cleaved into two fragments with distinct cytokine activities. The carboxy-terminal domain exhibits monocyte and leukocyte chemotaxis activity as well as stimulating production of myeloperoxidase, tumor necrosis factor- $\alpha$ , and tissue factor. The N-terminal domain binds to the interleukin-8 type A receptor and functions as an interleukin-8-like cytokine. Human tyrosyl-tRNA synthetase is secreted from  
 20 apoptotic tumor cells and may accelerate apoptosis (Wakasugi, K., and Schimmel, P. (1999) Science 284:147-151). Mitochondrial *Neurospora crassa* TyrRS and *S. cerevisiae* LeuRS are essential factors for certain group I intron splicing activities, and human mitochondrial LeuRS can substitute for the yeast LeuRS in a yeast null strain. Certain bacterial aaRSs are involved in regulating their own transcription or translation (Martinis, *supra*). Several aaRSs are able to synthesize diadenosine  
 25 oligophosphates, a class of signalling molecules with roles in cell proliferation, differentiation, and apoptosis (Kisselev, L.L et al. (1998) FEBS Lett. 427:157-163; Vartanian, A. et al. (1999) FEBS Lett. 456:175-180).

Autoantibodies against aminoacyl-tRNAs are generated by patients with autoimmune diseases such as rheumatic arthritis, dermatomyositis and polymyositis, and correlate strongly with complicating  
 30 interstitial lung disease (ILD) (Freist, W. et al. (1999) Biol. Chem. 380:623-646; Freist, W. et al. (1996) Biol. Chem. Hoppe Seyler 377:343-356). These antibodies appear to be generated in response to viral infection, and coxsackie virus has been used to induce experimental viral myositis in animals.

Comparison of aaRS structures between humans and pathogens has been useful in the design



of novel antibiotics (Schimmel, *supra*). Genetically engineered aaRSs have been utilized to allow site-specific incorporation of unnatural amino acids into proteins *in vivo* (Liu, D.R. et al. (1997) Proc. Natl. Acad. Sci. USA 94:10092-10097).

### tRNA Modifications

5           The modified ribonucleoside, pseudouridine ( $\psi$ ), is present ubiquitously in the anticodon regions of transfer RNAs (tRNAs), large and small ribosomal RNAs (rRNAs), and small nuclear RNAs (snRNAs).  $\psi$  is the most common of the modified nucleosides (i.e., other than G, A, U, and C) present in tRNAs. Only a few yeast tRNAs that are not involved in protein synthesis do not contain  $\psi$  (Cortese, R. et al. (1974) J. Biol. Chem. 249:1103-1108). The enzyme responsible for the conversion  
10 of uridine to  $\psi$ , pseudouridine synthase (pseudouridylylase), was first isolated from *Salmonella typhimurium* (Arena, F. et al. (1978) Nucleic Acids Res. 5:4523-4536). The enzyme has since been isolated from a number of mammals, including steer and mice (Green, C.J. et al. (1982) J. Biol. Chem. 257:3045-52; and Chen, J. and Patton, J.R. (1999) RNA 5:409-419). tRNA pseudouridine synthases have been the most extensively studied members of the family. They require a thiol donor (e.g.,  
15 cysteine) and a monovalent cation (e.g., ammonia or potassium) for optimal activity. Additional cofactors or high energy molecules (e.g., ATP or GTP) are not required (Green, *supra*). Other eukaryotic pseudouridine synthases have been identified that appear to be specific for rRNA (reviewed in Smith, C.M. and Steitz, J.A. (1997) Cell 89:669-672) and a dual-specificity enzyme has been identified that uses both tRNA and rRNA substrates (Wrzesinski, J. et al. (1995) RNA 1:  
20 437-448). The absence of  $\psi$  in the anticodon loop of tRNAs results in reduced growth in both bacteria (Singer, C.E. et al. (1972) Nature New Biol. 238:72-74) and yeast (Lecoite, F. (1998) J. Biol. Chem. 273:1316-1323), although the genetic defect is not lethal.

Another ribonucleoside modification that occurs primarily in eukaryotic cells is the conversion of guanosine to N<sup>2</sup>,N<sup>2</sup>-dimethylguanosine (m<sup>2</sup><sub>2</sub>G) at position 26 or 10 at the base of the D-stem of  
25 cytosolic and mitochondrial tRNAs. This posttranscriptional modification is believed to stabilize tRNA structure by preventing the formation of alternative tRNA secondary and tertiary structures. Yeast tRNA<sup>Asp</sup> is unusual in that it does not contain this modification. The modification does not occur in eubacteria, presumably because the structure of tRNAs in these cells and organelles is sequence constrained and does not require posttranscriptional modification to prevent the formation of  
30 alternative structures (Steinberg, S. and Cedergren, R. (1995) RNA 1:886-891, and references within). The enzyme responsible for the conversion of guanosine to m<sup>2</sup><sub>2</sub>G is a 63 kDa S-adenosylmethionine (SAM)-dependent tRNA N<sup>2</sup>,N<sup>2</sup>-dimethyl-guanosine methyltransferase (also referred to as the *TRM1* gene product and herein referred to as TRM) (Edqvist, J. (1995) Biochimie 77:54-61). The enzyme

localizes to both the nucleus and the mitochondria (Li, J-M. et al. (1989) J. Cell Biol. 109:1411-1419). Based on studies with TRM from *Xenopus laevis*, there appears to be a requirement for base pairing at positions C11-G24 and G10-C25 immediately preceding the G26 to be modified, with other structural features of the tRNA also being required for the proper presentation of the G26 substrate (Edqvist, J. et al. (1992) Nucleic Acids Res. 20:6575-6581). Studies in yeast suggest that cells carrying a weak ochre tRNA suppressor (*sup3-i*) are unable to suppress translation termination in the absence of TRM activity, suggesting a role for TRM in modifying the frequency of suppression in eukaryotic cells (Niederberger, C. et al. (1999) FEBS Lett. 464:67-70), in addition to the more general function of ensuring the proper three-dimensional structures for tRNA.

#### 10 Translation Initiation

Initiation of translation can be divided into three stages. The first stage brings an initiator transfer RNA (Met-tRNA<sub>i</sub>) together with the 40S ribosomal subunit to form the 43S preinitiation complex. The second stage binds the 43S preinitiation complex to the mRNA, followed by migration of the complex to the correct AUG initiation codon. The third stage brings the 60S ribosomal subunit to the 40S subunit to generate an 80S ribosome at the initiation codon. Regulation of translation primarily involves the first and second stage in the initiation process (V.M. Pain (1996) Eur. J. Biochem. 236:747-771).

Several initiation factors, many of which contain multiple subunits, are involved in bringing an initiator tRNA and the 40S ribosomal subunit together. eIF2, a guanine nucleotide binding protein, recruits the initiator tRNA to the 40S ribosomal subunit. Only when eIF2 is bound to GTP does it associate with the initiator tRNA. eIF2B, a guanine nucleotide exchange protein, is responsible for converting eIF2 from the GDP-bound inactive form to the GTP-bound active form. Two other factors, eIF1A and eIF3 bind and stabilize the 40S subunit by interacting with the 18S ribosomal RNA and specific ribosomal structural proteins. eIF3 is also involved in association of the 40S ribosomal subunit with mRNA. The Met-tRNA<sub>i</sub>, eIF1A, eIF3, and 40S ribosomal subunit together make up the 43S preinitiation complex (Pain, *supra*).

Additional factors are required for binding of the 43S preinitiation complex to an mRNA molecule, and the process is regulated at several levels. eIF4F is a complex consisting of three proteins: eIF4E, eIF4A, and eIF4G. eIF4E recognizes and binds to the mRNA 5'-terminal m<sup>7</sup>GTP cap, eIF4A is a bidirectional RNA-dependent helicase, and eIF4G is a scaffolding polypeptide. eIF4G has three binding domains. The N-terminal third of eIF4G interacts with eIF4E, the central third interacts with eIF4A, and the C-terminal third interacts with eIF3 bound to the 43S preinitiation complex. Thus, eIF4G acts as a bridge between the 40S ribosomal subunit and the mRNA (M.W.

Hentze (1997) Science 275:500-501).

The ability of eIF4F to initiate binding of the 43S preinitiation complex is regulated by structural features of the mRNA. The mRNA molecule has an untranslated region (UTR) between the 5' cap and the AUG start codon. In some mRNAs this region forms secondary structures that  
5 impede binding of the 43S preinitiation complex. The helicase activity of eIF4A is thought to function in removing this secondary structure to facilitate binding of the 43S preinitiation complex (Pain, *supra*).

### **Translation Elongation**

Elongation is the process whereby additional amino acids are joined to the initiator methionine  
10 to form the complete polypeptide chain. The elongation factors EF1  $\alpha$ , EF1  $\beta$   $\gamma$ , and EF2 are involved in elongating the polypeptide chain following initiation. EF1  $\alpha$  is a GTP-binding protein. In EF1  $\alpha$ 's GTP-bound form, it brings an aminoacyl-tRNA to the ribosome's A site. The amino acid attached to the newly arrived aminoacyl-tRNA forms a peptide bond with the initiator methionine. The GTP on EF1  $\alpha$  is hydrolyzed to GDP, and EF1  $\alpha$ -GDP dissociates from the ribosome. EF1  $\beta$   $\gamma$  binds EF1  $\alpha$ -  
15 GDP and induces the dissociation of GDP from EF1  $\alpha$ , allowing EF1  $\alpha$  to bind GTP and a new cycle to begin.

As subsequent aminoacyl-tRNAs are brought to the ribosome, EF-G, another GTP-binding protein, catalyzes the translocation of tRNAs from the A site to the P site and finally to the E site of the ribosome. This allows the ribosome and the mRNA to remain attached during translation.

### **Translation Termination**

The release factor eRF carries out termination of translation. eRF recognizes stop codons in the mRNA, leading to the release of the polypeptide chain from the ribosome.

### **Expression profiling**

Microarrays are analytical tools used in bioanalysis. A microarray has a plurality of molecules  
25 spatially distributed over, and stably associated with, the surface of a solid support. Microarrays of polypeptides, polynucleotides, and/or antibodies have been developed and find use in a variety of applications, such as gene sequencing, monitoring gene expression, gene mapping, bacterial identification, drug discovery, and combinatorial chemistry.

One area in particular in which microarrays find use is in gene expression analysis. Array  
30 technology can provide a simple way to explore the expression of a single polymorphic gene or the expression profile of a large number of related or unrelated genes. When the expression of a single gene is examined, arrays are employed to detect the expression of a specific gene or its variants. When an expression profile is examined, arrays provide a platform for identifying genes that are tissue

specific, are affected by a substance being tested in a toxicology assay, are part of a signaling cascade, carry out housekeeping functions, or are specifically related to a particular genetic predisposition, condition, disease, or disorder.

#### RNA Expression

5        Array technology can provide a simple way to explore the expression of a single polymorphic gene or the expression profile of a large number of related or unrelated genes. When the expression of a single gene is examined, arrays are employed to detect the expression of a specific gene or its variants. When an expression profile is examined, arrays provide a platform for examining which genes are tissue specific, carrying out housekeeping functions, parts of a signaling cascade, or  
10    specifically related to a particular genetic predisposition, condition, disease, or disorder.

      The potential application of gene expression profiling is particularly relevant to measuring the toxic response to potential therapeutic compounds and of the metabolic response to therapeutic agents. Diseases treated with steroids and disorders caused by the metabolic response to treatment with steroids include adenomatosis, cholestasis, cirrhosis, hemangioma, Henoch-Schonlein purpura,  
15    hepatitis, hepatocellular and metastatic carcinomas, idiopathic thrombocytopenic purpura, porphyria, sarcoidosis, and Wilson disease. Response may be measured by comparing both the levels and sequences expressed in tissues from subjects exposed to or treated with steroid compounds such as mifepristone, progesterone, beclomethasone, medroxyprogesterone, budesonide, prednisone, dexamethasone, betamethasone, or danazol with the levels and sequences expressed in normal  
20    untreated tissue.

      Steroids are a class of lipid-soluble molecules, including cholesterol, bile acids, vitamin D, and hormones, that share a common four-ring structure based on cyclopentanoperhydrophenanthrene and that carry out a wide variety of functions. Cholesterol, for example, is a component of cell membranes that controls membrane fluidity. It is also a precursor for bile acids which solubilize lipids  
25    and facilitate absorption in the small intestine during digestion. Vitamin D regulates the absorption of calcium in the small intestine and controls the concentration of calcium in plasma. Steroid hormones, produced by the adrenal cortex, ovaries, and testes, include glucocorticoids, mineralocorticoids, androgens, and estrogens. They control various biological processes by binding to intracellular receptors that regulate transcription of specific genes in the nucleus. Glucocorticoids, for example,  
30    increase blood glucose concentrations by regulation of gluconeogenesis in the liver, increase blood concentrations of fatty acids by promoting lipolysis in adipose tissues, modulate sensitivity to catecholamines in the central nervous system, and reduce inflammation. The principal

mineralocorticoid, aldosterone, is produced by the adrenal cortex and acts on cells of the distal tubules of the kidney to enhance sodium ion reabsorption. Androgens, produced by the interstitial cells of Leydig in the testis, include the male sex hormone testosterone, which triggers changes at puberty, the production of sperm and maintenance of secondary sexual characteristics. Female sex hormones, estrogen and progesterone, are produced by the ovaries and also by the placenta and adrenal cortex of the fetus during pregnancy. Estrogen regulates female reproductive processes and secondary sexual characteristics. Progesterone regulates changes in the endometrium during the menstrual cycle and pregnancy.

Steroid hormones are widely used for fertility control and in anti-inflammatory treatments for physical injuries and diseases such as arthritis, asthma, and auto-immune disorders. Progesterone, a naturally occurring progestin, is primarily used to treat amenorrhea, abnormal uterine bleeding, or as a contraceptive. Endogenous progesterone is responsible for inducing secretory activity in the endometrium of the estrogen-primed uterus in preparation for the implantation of a fertilized egg and for the maintenance of pregnancy. It is secreted from the corpus luteum in response to luteinizing hormone (LH). The primary contraceptive effect of exogenous progestins involves the suppression of the midcycle surge of LH. At the cellular level, progestins diffuse freely into target cells and bind to the progesterone receptor. Target cells include the female reproductive tract, the mammary gland, the hypothalamus, and the pituitary. Once bound to the receptor, progestins slow the frequency of release of gonadotropin releasing hormone from the hypothalamus and blunt the pre-ovulatory LH surge, thereby preventing follicular maturation and ovulation. Progesterone has minimal estrogenic and androgenic activity. Progesterone is metabolized hepatically to pregnanediol and conjugated with glucuronic acid.

Medroxyprogesterone (MAH), also known as 6 $\alpha$ -methyl-17-hydroxyprogesterone, is a synthetic progestin with a pharmacological activity about 15 times greater than progesterone. MAH is used for the treatment of renal and endometrial carcinomas, amenorrhea, abnormal uterine bleeding, and endometriosis associated with hormonal imbalance. MAH has a stimulatory effect on respiratory centers and has been used in cases of low blood oxygenation caused by sleep apnea, chronic obstructive pulmonary disease, or hypercapnia.

Mifepristone, also known as RU-486, is an antiprogesterone drug that blocks receptors of progesterone. It counteracts the effects of progesterone, which is needed to sustain pregnancy. Mifepristone induces spontaneous abortion when administered in early pregnancy followed by treatment with the prostaglandin, misoprostol. Further, studies show that mifepristone at a

substantially lower dose can be highly effective as a postcoital contraceptive when administered within five days after unprotected intercourse, thus providing women with a "morning-after pill" in case of contraceptive failure or sexual assault. Mifepristone also has potential uses in the treatment of breast and ovarian cancers in cases in which tumors are progesterone-dependent. It interferes with steroid-dependent growth of brain meningiomas, and may be useful in treatment of endometriosis where it blocks the estrogen-dependent growth of endometrial tissues. It may also be useful in treatment of uterine fibroid tumors and Cushing's Syndrome. Mifepristone binds to glucocorticoid receptors and interferes with cortisol binding. Mifepristone also may act as an anti-glucocorticoid and be effective for treating conditions where cortisol levels are elevated such as AIDS, anorexia nervosa, ulcers, diabetes, Parkinson's disease, multiple sclerosis, and Alzheimer's disease.

Danazol is a synthetic steroid derived from ethinyl testosterone. Danazol indirectly reduces estrogen production by lowering pituitary synthesis of follicle-stimulating hormone and LH. Danazol also binds to sex hormone receptors in target tissues, thereby exhibiting anabolic, antiestrogenic, and weakly androgenic activity. Danazol does not possess any progestogenic activity, and does not suppress normal pituitary release of corticotropin or release of cortisol by the adrenal glands. Danazol is used in the treatment of endometriosis to relieve pain and inhibit endometrial cell growth. It is also used to treat fibrocystic breast disease and hereditary angioedema.

Corticosteroids are used to relieve inflammation and to suppress the immune response. They inhibit eosinophil, basophil, and airway epithelial cell function by regulation of cytokines that mediate the inflammatory response. They inhibit leukocyte infiltration at the site of inflammation, interfere in the function of mediators of the inflammatory response, and suppress the humoral immune response. Corticosteroids are used to treat allergies, asthma, arthritis, and skin conditions. Beclomethasone is a synthetic glucocorticoid that is used to treat steroid-dependent asthma, to relieve symptoms associated with allergic or nonallergic (vasomotor) rhinitis, or to prevent recurrent nasal polyps following surgical removal. The anti-inflammatory and vasoconstrictive effects of intranasal beclomethasone are 5000 times greater than those produced by hydrocortisone. Budesonide is a corticosteroid used to control symptoms associated with allergic rhinitis or asthma. Budesonide has high topical anti-inflammatory activity but low systemic activity. Dexamethasone is a synthetic glucocorticoid used in anti-inflammatory or immunosuppressive compositions. It is also used in inhalants to prevent symptoms of asthma. Due to its greater ability to reach the central nervous system, dexamethasone is usually the treatment of choice to control cerebral edema. Dexamethasone is approximately 20-30 times more potent than hydrocortisone and 5-7 times more potent than prednisone. Prednisone is metabolized in

the liver to its active form, prednisolone, a glucocorticoid with anti-inflammatory properties.

Prednisone is approximately 4 times more potent than hydrocortisone and the duration of action of prednisone is intermediate between hydrocortisone and dexamethasone. Prednisone is used to treat allograft rejection, asthma, systemic lupus erythematosus, arthritis, ulcerative colitis, and other

5 inflammatory conditions. Betamethasone is a synthetic glucocorticoid with antiinflammatory and immunosuppressive activity and is used to treat psoriasis and fungal infections, such as athlete's foot and ringworm.

The anti-inflammatory actions of corticosteroids are thought to involve phospholipase A<sub>2</sub> inhibitory proteins, collectively called lipocortins. Lipocortins, in turn, control the biosynthesis of potent  
10 mediators of inflammation such as prostaglandins and leukotrienes by inhibiting the release of the precursor molecule arachidonic acid. Proposed mechanisms of action include decreased IgE synthesis, increased number of  $\beta$ -adrenergic receptors on leukocytes, and decreased arachidonic acid metabolism. During an immediate allergic reaction, such as in chronic bronchial asthma, allergens bridge the IgE antibodies on the surface of mast cells, which triggers these cells to release  
15 chemotactic substances. Mast cell influx and activation, therefore, is partially responsible for the inflammation and hyperirritability of the oral mucosa in asthmatic patients. This inflammation can be retarded by administration of corticosteroids.

#### CA3+ Steroids

Array technology can provide a simple way to explore the expression of a single polymorphic  
20 gene or the expression profile of a large number of related or unrelated genes. When the expression of a single gene is examined, arrays are employed to detect the expression of a specific gene or its variants. When an expression profile is examined, arrays provide a platform for examining which genes are tissue specific, carrying out housekeeping functions, parts of a signaling cascade, or specifically related to a particular genetic predisposition, condition, disease, or disorder.

25 The potential application of gene expression profiling is particularly relevant to measuring the toxic response to potential therapeutic compounds and of the metabolic response to therapeutic agents. Diseases treated with steroids and disorders caused by the metabolic response to treatment with steroids include adenomatosis, cholestasis, cirrhosis, hemangioma, Henoch-Schonlein purpura, hepatitis, hepatocellular and metastatic carcinomas, idiopathic thrombocytopenic purpura, porphyria,  
30 sarcoidosis, and Wilson disease. Response may be measured by comparing both the levels and sequences expressed in tissues from subjects exposed to or treated with steroid compounds such as mifepristone, progesterone, beclomethasone, medroxyprogesterone, budesonide, prednisone,

dexamethasone, betamethasone, or danazol with the levels and sequences expressed in normal untreated tissue.

The effects upon liver metabolism and hormone clearance mechanisms are important to understand the pharmacodynamics of a drug. The human C3A cell line is a clonal derivative of HepG2/C3 (hepatoma cell line, isolated from a 15-year-old male with liver tumor), which was selected for strong contact inhibition of growth. The use of a clonal population enhances the reproducibility of the cells. C3A cells have many characteristics of primary human hepatocytes in culture: i) expression of insulin receptor and insulin-like growth factor II receptor; ii) secretion of a high ratio of serum albumin compared with  $\alpha$ -fetoprotein iii) conversion of ammonia to urea and glutamine; iv) metabolize aromatic amino acids; and v) proliferate in glucose-free and insulin-free medium. The C3A cell line is now well established as an in vitro model of the mature human liver (Mickelson et al. (1995) Hepatology 22:866-875; Nagendra et al. (1997) Am J Physiol 272:G408-G416).

#### Vascular endothelial activation

Activation of vascular endothelium is a central event in a wide range of physiological and disease processes such as vascular tone regulation, coagulation and thrombosis, atherosclerosis, inflammation and some infectious diseases. Blood vessel walls are composed of two tissue layers: an endothelial cell (EC) layer which comprises the luminal surface of the vessel, and an underlying vascular smooth muscle cell (VSMC) layer. Through dynamic interactions with each other and with surrounding tissues, the vascular endothelium and smooth muscle tissues maintain vascular tone, control selective permeability of the vascular wall, direct vessel remodeling and angiogenesis, and modulate inflammatory and immune responses.

The inflammatory response is a complex vascular reaction mediated by numerous cytokines, chemokines, growth factors, and other signaling molecules expressed by activated ECs, VSMCs and leukocytes. Inflammation protects the organism during trauma and infection, but can also lead to pathological conditions such as atherosclerosis.

Human coronary artery endothelial cells (HCAECs) are primary cells derived from the endothelium of a human coronary artery. HCAECs are used as an experimental model for investigating the role of the endothelium in human vascular biology in vitro. Human umbilical artery endothelial cells (HUAECs) are primary cells derived from the endothelium of an umbilical artery. Human uterine myometrium microvascular endothelial cells (UtMVECs) are primary cells derived from the uterine myometrium microvasculature. Human Iliac Artery Endothelial Cells (HIAECs) are primary cells derived from the endothelium of an iliac artery. Human umbilical vein endothelial cells



(HUVECs) are a primary cell line derived from the endothelium of the human umbilical vein. ECV304 is a human endothelial line.

#### Breast cancer

There are more than 180,000 new cases of breast cancer diagnosed each year, and the mortality rate for breast cancer approaches 10% of all deaths in females between the ages of 45-54 (Gish, K. (1999) AWIS Magazine 28:7-10). However the survival rate based on early diagnosis of localized breast cancer is extremely high (97%), compared with the advanced stage of the disease in which the tumor has spread beyond the breast (22%). Current procedures for clinical breast examination are lacking in sensitivity and specificity, and efforts are underway to develop comprehensive gene expression profiles for breast cancer that may be used in conjunction with conventional screening methods to improve diagnosis and prognosis of this disease (Perou, C.M. et al. (2000) Nature 406:747-752). Mutations in two genes, BRCA1 and BRCA2, are known to greatly predispose a woman to breast cancer and may be passed on from parents to children (Gish, supra). However, this type of hereditary breast cancer accounts for only about 5% to 9% of breast cancers, while the vast majority of breast cancer is due to non-inherited mutations that occur in breast epithelial cells.

Cell lines derived from human mammary epithelial cells at various stages of breast cancer provide a useful model to study the process of malignant transformation and tumor progression as it has been shown that these cell lines retain many of the properties of their parental tumors for lengthy culture periods (Wistuba, I.I. et al. (1998) Clin. Cancer Res. 4:2931-2938). Such a model is particularly useful for comparing phenotypic and molecular characteristics of human mammary epithelial cells at various stages of malignant transformation.

Estrogen stimulation plays a critical role in the development of normal mammary epithelium. Estradiol has a direct mitogenic effect on breast cancer cells, causing them to divide more rapidly by shortening their cell cycle. Also, estradiol induces a large number of enzymes and other proteins involved in nucleic acid synthesis in isolated breast cancer cell lines. Estradiol may increase the expression of the EGF receptor in response to TGF- $\alpha$  and EGF. In addition, estrogens may promote proliferation of tumor cells by inducing the synthesis of TGF- $\alpha$  and EGF, and may block growth factors that would normally inhibit tumor cell growth. Estrogen receptor (ER) has been investigated extensively as a prognostic marker in breast cancer. Patients whose tumors display high levels of estrogen receptor have a significantly better prognosis than patients with receptor-negative tumors. When ER is lost or cells expressing the ER are selected against by therapeutic treatments, the tumor

becomes more aggressive.

#### Lung cancer

Lung cancer is the leading cause of cancer death for men and the second leading cause of cancer death for women in the U.S. Lung cancers are divided into four histopathologically distinct groups. Three groups (squamous cell carcinoma, adenocarcinoma, and large cell carcinoma) are classified as non-small cell lung cancers (NSCLCs). The fourth group of cancers is referred to as small cell lung cancer (SCLC). Deletions on chromosome 3 are common in this disease and are thought to indicate the presence of a tumor suppressor gene in this region. Activating mutations in K-ras are commonly found in lung cancer and are the basis of one of the mouse models for the disease.

#### 10 Ovarian Cancer

Ovarian cancer is the leading cause of death from a gynecologic cancer. The majority of ovarian cancers are derived from epithelial cells, and 70% of patients with epithelial ovarian cancers present with late-stage disease. As a result, the long-term survival rates for this disease is very low. Identification of early-stage markers for ovarian cancer would significantly increase the survival rate.

15 The molecular events that lead to ovarian cancer are poorly understood. Some of the known aberrations include mutation of p53 and microsatellite instability. Since gene expression patterns are likely to vary when normal ovary is compared to ovarian tumors, examination of gene expression in these tissues to identify possible markers for ovarian cancer is particularly relevant to improving diagnosis, prognosis, and treatment of this disease.

#### 20 Colon Cancer

Colorectal cancer is the second leading cause of cancer deaths in the United States. Colon cancer is associated with aging, since 90% of the total cases occur in individuals over the age of 55. A widely accepted hypothesis is that several contributing genetic mutations must accumulate over time in an individual who develops the disease. To understand the nature of genetic alterations in colorectal cancer, a number of studies have focused on the inherited syndromes. The first known inherited syndrome, Familial Adenomatous Polyposis (FAP), is caused by mutations in the Adenomatous Polyposis Coli gene (APC), resulting in truncated or inactive forms of the protein. This tumor suppressor gene has been mapped to chromosome 5q. The second known inherited syndrome is hereditary nonpolyposis colorectal cancer (HNPCC), which is caused by mutations in mismatch repair genes.

30

Although hereditary colon cancer syndromes occur in a small percentage of the population and most colorectal cancers are considered sporadic, knowledge from studies of the hereditary

syndromes can be generally applied. For instance, somatic mutations in APC occur in at least 80% of indiscriminate colon tumors. APC mutations are thought to be the initiating event in the disease. Other mutations occur subsequently. Approximately 50% of colorectal cancers contain activating mutations in ras, while 85% contain inactivating mutations in p53. Changes in these genes lead to gene expression changes in colon cancer. Less is understood about downstream targets of these mutations and the role they may play in cancer development and progression.

#### Inflammatory/Immune Response

Blood vessel walls are composed of two tissue layers: an endothelial cell (EC) layer which comprises the luminal surface of the vessel, and an underlying vascular smooth muscle cell (VSMC) layer. Through dynamic interactions with each other and with surrounding tissues, the vascular endothelium and smooth muscle tissues maintain vascular tone, control selective permeability of the vascular wall, direct vessel remodeling and angiogenesis, and modulate inflammatory and immune responses.

The inflammatory response is a complex vascular reaction mediated by numerous cytokines, chemokines, growth factors, and other signaling molecules expressed by activated ECs, VSMCs and leukocytes. Inflammation protects the organism during trauma and infection, but can also lead to pathological conditions such as atherosclerosis. The pro-inflammatory cytokines, interleukin (IL)-1 and tumor necrosis factor (TNF), are secreted by a small number of activated macrophages or other cells and can set off a cascade of vascular changes, largely through their ability to alter gene expression patterns in ECs and VSMCs. These vascular changes include vasodilation and increased permeability of microvasculature, edema, and leukocyte extravasation and transmigration across the vessel wall. Ultimately, leukocytes, particularly neutrophils and monocytes/macrophages, accumulate in the extravascular space, where they remove injurious agents by phagocytosis and oxidative killing, a process accompanied by release of toxic factors, such as proteases and reactive oxygen species.

Tumor necrosis factor- $\alpha$  (TNF- $\alpha$ ) is a pleiotropic factor that exerts a variety of effects, such as growth promotion, growth inhibition, angiogenesis, cytotoxicity, inflammation, and immunomodulation. This cytokine is synthesized mainly by macrophages in response to invasive stimuli as an active 26 kDa membrane-bound precursor that is cleaved proteolytically to a mature 17 kDa form with the prosequence polypeptide remaining associated to the membrane. The peptide is bioactive as a 51 kDa trimer, which can be recognized by TNF- $\alpha$  receptors. TNF- $\alpha$  receptors are present in the majority of cell types. IL-1 and TNF induce pro-inflammatory, thrombotic, and anti-apoptotic changes in gene expression by signaling through receptors on the surface of ECs and

VSMCs; these receptors activate transcription factors such as NF $\kappa$ B as well as AP-1, IRF-1, and NF-GMa, leading to alterations in gene expression. Genes known to be differentially regulated in EC by IL-1 and TNF include E selectin, VCAM-1, ICAM-1, PAF, I $\kappa$ B $\alpha$ , IAP-1, MCP-1, eotaxin, ENA-78, G-CSF, A20, ICE, and complement C3 component. A key event in inflammation, adhesion and transmigration of blood leukocytes across the vascular endothelium, for example, is mediated by increased expression of E selectin, P selectin, ICAM-1, and VCAM-1 on activated endothelium.

Several investigators have examined changes in vascular cell gene expression associated with various inflammatory diseases or model systems. Examining human umbilical vein endothelial cells (HUVEC) activated by recombinant TNF $\alpha$  or conditioned medium from activated human primary monocytes, Horrevoets *et al.* (1999; Blood 93:3418-3431) identified 106 differentially regulated genes. In a similar approach, deVries *et al.* (2000; JBC 275:23939-23947) identified 40 differentially regulated genes in umbilical cord artery-derived smooth muscle cells activated by conditioned media from cultured macrophages after stimulation with oxidized LDL particles. In both studies, many of the identified genes were already known to be involved in inflammation. Comparing expression profiles from inflammatory diseased tissues, cultured macrophages, chondrocyte cell lines, primary chondrocytes, and synoviocytes, Heller *et al.* (1997; Proc Natl Acad Sci USA 94:2150-2155) identified candidate genes involved in inflammatory responses, including TNF, IL-1 IL-6, IL-8 G-CSF, RANTES, and V-CAM. From this candidate gene set, tissue inhibitor of metalloproteinase 1, ferritin light chain, and manganese superoxide dismutase were found to be differentially expressed in rheumatoid arthritis (RA) relative to inflammatory bowel disease (IBD). Further, IL-3, chemokine Gro $\alpha$ , and metalloproteinase matrix metallo-elastase were expressed in both RA and IBD. Most recently, in an analysis of cultured aortic smooth muscle cells treated with TNF $\alpha$ , Haley *et al.* (2000; Circulation 102:2185-2189) found a 20-fold increase in eotaxin, an eosinophil chemotactic factor. The overexpression of eotaxin and its receptor CCR3 in atherosclerotic lesions was confirmed by northern analysis.

Array technology can provide a simple way to explore the expression of a single polymorphic gene or the expression profile of a large number of related or unrelated genes. When the expression of a single gene is examined, arrays are employed to detect the expression of a specific gene or its variants. When an expression profile is examined, arrays provide a platform for identifying genes that are tissue specific, carry out housekeeping functions, are part of a signaling cascade, or are specifically related to a particular genetic predisposition, condition, disease, or disorder. The potential application of gene expression profiling is particularly relevant to improving diagnosis, prognosis, and

treatment of disease. For example, both the levels and sequences expressed in tissues from subjects with coronary artery disease may be compared with the levels and sequences expressed in normal vascular tissue.

DNA-based array technology is also useful as a method of analyzing cell signaling pathways.

- 5 For example, protein kinase C (PKC) is a family of serine/threonine kinases which play a critical role in many signal transduction pathways in the cell (Kanashiro, C.A. and Khalil, R.A. (1998) Clin. Exp. Pharmacol. Physiol. 25:974-985). Phorbol 12-myristate 13-acetate (PMA) is a broad activator of the protein kinase C-dependent pathways. PMA promotes tumors in cells by over activating PKC pathways. PKC is also affected by intracellular calcium levels. Internal calcium flux has been shown  
10 have many effects on cells, particular in the areas of cell activation and proliferation (Cole, K, and Kohn, E. (1994) Cancer Metastasis Rev. 13:31-44). Ionomycin is a calcium ionophore that permits the entry of calcium into the cell, hence increasing the cytosolic calcium concentration. Thus the combination of PMA and ionomycin activates two of the major signaling pathways used by mammalian cells to interact with their environment. In T cells, for example, the combination of PMA  
15 and ionomycin mimics the type of secondary signaling events elicited during optimal T cell activation, while in endothelial cells, PMA and ionomycin reproduce the effect of inflammatory mediators.

- Cell lines are widely used in experimental biology to model human cell behavior. Jurkat, an acute T-cell leukemia cell line that grows actively in the absence of external stimuli, is used to study signaling in human T cells. ECV304, a cell line derived from the endothelium of the human umbilical  
20 vein, is used to study the functional biology of human endothelial cells.

#### Neurological disorders

- Characterization of region-specific gene expression in the human brain provides a context and background for molecular neurobiology on a variety of neurological disorders. For example, Alzheimer's disease (AD) is a progressive, neurodestructive process of the human neocortex,  
25 characterized by the deterioration of memory and higher cognitive function. A progressive and irreversible brain disorder, AD is characterized by three major pathogenic episodes involving (a) an aberrant processing and deposition of beta-amyloid precursor protein (betaAPP) to form neurotoxic beta-amyloid (betaA) peptides and an aggregated insoluble polymer of betaA that forms the senile plaque, (b) the establishment of intraneuronal neuritic tau pathology yielding widespread deposits of  
30 agyrophilic neurofibrillary tangles (NFT) and (c) the initiation and proliferation of a brain-specific inflammatory response. These three seemingly disperse attributes of AD etiopathogenesis are linked by the fact that proinflammatory microglia, reactive astrocytes and their associated cytokines and

chemokines are associated with the biology of the microtubule associated protein tau, betaA speciation and aggregation. Missense mutations in the presenilin genes PS1 and PS2, implicated in early onset familial AD, cause abnormal betaAPP processing with resultant overproduction of betaA42 and related neurotoxic peptides. Specific betaA fragments such as betaA42 can further potentiate proinflammatory mechanisms. Expression of the inducible oxidoreductase cyclooxygenase-2 and cytosolic phospholipase A2 (cPLA2) are strongly activated during cerebral ischemia and trauma, epilepsy and AD, indicating the induction of proinflammatory gene pathways as a response to brain injury. Neurotoxic metals such as aluminum and zinc, both implicated in AD etiopathogenesis, and arachidonic acid, a major metabolite of brain cPLA2 activity, each polymerize hyperphosphorylated tau to form NFT-like bundles. Studies have identified a reduced risk for AD in patients aged over 70 years who were previously treated with non-steroidal anti-inflammatory drugs for non-CNS afflictions that include arthritis. (For a review of the interrelationships between the mechanisms of PS1, PS2 and betaAPP gene expression, tau and betaA deposition and the induction, regulation and proliferation in AD of the neuroinflammatory response, see Lukiw W.J, and Bazan N.G.(2000) Neurochem. Res. 2000 25:1173-1184).

The discovery of new nucleic acid-associated proteins, and the polynucleotides encoding them, satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of cell proliferative, neurological, developmental, and autoimmune/inflammatory disorders, and infections, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of nucleic acid-associated proteins.

There is a need in the art for new compositions, including nucleic acids and proteins, for the diagnosis, prevention, and treatment of cell proliferative, neurological, developmental, and autoimmune/inflammatory disorders, and infections.

## SUMMARY OF THE INVENTION

Various embodiments of the invention provide purified polypeptides, nucleic acid-associated proteins, referred to collectively as 'NAAP' and individually as 'NAAP-1,' 'NAAP-2,' 'NAAP-3,' 'NAAP-4,' 'NAAP-5,' 'NAAP-6,' 'NAAP-7,' 'NAAP-8,' 'NAAP-9,' 'NAAP-10,' 'NAAP-11,' 'NAAP-12,' 'NAAP-13,' 'NAAP-14,' 'NAAP-15,' 'NAAP-16,' 'NAAP-17,' 'NAAP-18,' 'NAAP-19,' 'NAAP-20,' 'NAAP-21,' 'NAAP-22,' 'NAAP-23,' 'NAAP-24,' 'NAAP-25,' 'NAAP-26,' 'NAAP-27,' 'NAAP-28,' 'NAAP-29,' 'NAAP-30,' 'NAAP-31,' 'NAAP-32,' 'NAAP-33,' 'NAAP-34,' 'NAAP-35,' 'NAAP-36,' 'NAAP-37,' 'NAAP-38,' 'NAAP-39,' 'NAAP-40,' 'NAAP-41,'

'NAAP-42,' 'NAAP-43,' 'NAAP-44,' 'NAAP-45,' 'NAAP-46,' 'NAAP-47,' 'NAAP-48,' 'NAAP-49,' 'NAAP-50,' 'NAAP-51,' 'NAAP-52,' 'NAAP-53,' 'NAAP-54,' 'NAAP-55,' 'NAAP-56,' and 'NAAP-57' and methods for using these proteins and their encoding polynucleotides for the detection, diagnosis, and treatment of diseases and medical conditions. Embodiments also provide methods for  
5 utilizing the purified nucleic acid-associated proteins and/or their encoding polynucleotides for facilitating the drug discovery process, including determination of efficacy, dosage, toxicity, and pharmacology. Related embodiments provide methods for utilizing the purified nucleic acid-associated proteins and/or their encoding polynucleotides for investigating the pathogenesis of diseases and medical conditions.

- 10 An embodiment provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-57, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-57, c) a biologically active fragment of a polypeptide having an amino acid sequence selected  
15 from the group consisting of SEQ ID NO:1-57, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-57. Another embodiment provides an isolated polypeptide comprising an amino acid sequence of SEQ ID NO:1-57.

- Still another embodiment provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-57, b) a polypeptide comprising a naturally occurring amino acid  
20 sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-57, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-57, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-57. In another embodiment, the polynucleotide encodes a polypeptide  
25 selected from the group consisting of SEQ ID NO:1-57. In an alternative embodiment, the polynucleotide is selected from the group consisting of SEQ ID NO:58-114.

- Still another embodiment provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group  
30 consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-57, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group

consisting of SEQ ID NO:1-57, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-57, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-57. Another embodiment provides a cell transformed with the recombinant polynucleotide. Yet another  
5 embodiment provides a transgenic organism comprising the recombinant polynucleotide.

Another embodiment provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-57, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group  
10 consisting of SEQ ID NO:1-57, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-57, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-57. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence  
15 operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Yet another embodiment provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-57, b) a polypeptide comprising a naturally  
20 occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-57, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-57, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-57.

25 Still yet another embodiment provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:58-114, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:58-114, c) a polynucleotide complementary to the  
30 polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In other embodiments, the polynucleotide can comprise at least about 20, 30, 40, 60, 80, or 100 contiguous nucleotides.



Yet another embodiment provides a method for detecting a target polynucleotide in a sample, said target polynucleotide being selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:58-114, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:58-114, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex. In a related embodiment, the method can include detecting the amount of the hybridization complex. In still other embodiments, the probe can comprise at least about 20, 30, 40, 60, 80, or 100 contiguous nucleotides.

Still yet another embodiment provides a method for detecting a target polynucleotide in a sample, said target polynucleotide being selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:58-114, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:58-114, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof. In a related embodiment, the method can include detecting the amount of the amplified target polynucleotide or fragment thereof.

Another embodiment provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-57, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-57, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-57, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the

group consisting of SEQ ID NO:1-57, and a pharmaceutically acceptable excipient. In one embodiment, the composition can comprise an amino acid sequence selected from the group consisting of SEQ ID NO:1-57. Other embodiments provide a method of treating a disease or condition associated with decreased or abnormal expression of functional NAAP, comprising administering to a patient in need of such treatment the composition.

Yet another embodiment provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-57, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-57, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-57, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-57. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. Another embodiment provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. Yet another embodiment provides a method of treating a disease or condition associated with decreased expression of functional NAAP, comprising administering to a patient in need of such treatment the composition.

Still yet another embodiment provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-57, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-57, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-57, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-57. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. Another embodiment provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. Yet another embodiment provides a method of treating a disease or condition associated with overexpression of functional NAAP, comprising administering to a patient in need of such treatment the composition.

Another embodiment provides a method of screening for a compound that specifically binds to

a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-57, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-57, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-57, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-57. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

Yet another embodiment provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-57, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-57, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-57, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-57. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

Still yet another embodiment provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:58-114, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, b) detecting altered expression of the target polynucleotide, and c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

Another embodiment provides a method for assessing toxicity of a test compound, said

method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:58-114, ii) a  
 5 polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:58-114, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target  
 10 polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:58-114, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:58-114, iii) a polynucleotide complementary to the polynucleotide of i), iv) a  
 15 polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide can comprise a fragment of a polynucleotide selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization  
 20 complex in the treated biological sample is indicative of toxicity of the test compound.

### BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for full length polynucleotide and polypeptide embodiments of the invention.

25 Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog, and the PROTEOME database identification numbers and annotations of PROTEOME database homologs, for polypeptide embodiments of the invention. The probability scores for the matches between each polypeptide and its homolog(s) are also shown.

Table 3 shows structural features of polypeptide embodiments, including predicted motifs and  
 30 domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble

polynucleotide embodiments, along with selected fragments of the polynucleotides.

Table 5 shows representative cDNA libraries for polynucleotide embodiments.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

5 Table 7 shows the tools, programs, and algorithms used to analyze polynucleotides and polypeptides, along with applicable descriptions, references, and threshold parameters.

Table 8 shows single nucleotide polymorphisms found in polynucleotide sequences of the invention, along with allele frequencies in different human populations.

## 10 DESCRIPTION OF THE INVENTION

Before the present proteins, nucleic acids, and methods are described, it is understood that embodiments of the invention are not limited to the particular machines, instruments, materials, and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the  
15 invention.

As used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

20 Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the  
25 cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with various embodiments of the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

## DEFINITIONS

"NAAP" refers to the amino acid sequences of substantially purified NAAP obtained from  
30 any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of

NAAP. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of NAAP either by directly interacting with NAAP or by acting on components of the biological pathway in which NAAP participates.

An "allelic variant" is an alternative form of the gene encoding NAAP. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding NAAP include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as NAAP or a polypeptide with at least one functional characteristic of NAAP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding NAAP, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide encoding NAAP. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent NAAP. Deliberate amino acid substitutions may be made on the basis of one or more similarities in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of NAAP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" can refer to an oligopeptide, a peptide, a polypeptide, or a protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid. Amplification

may be carried out using polymerase chain reaction (PCR) technologies or other nucleic acid amplification technologies well known in the art.

The term “antagonist” refers to a molecule which inhibits or attenuates the biological activity of NAAP. Antagonists may include proteins such as antibodies, anticalins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of NAAP either by directly interacting with NAAP or by acting on components of the biological pathway in which NAAP participates.

The term “antibody” refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')<sub>2</sub>, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind NAAP polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term “antigenic determinant” refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term “aptamer” refers to a nucleic acid or oligonucleotide molecule that binds to a specific molecular target. Aptamers are derived from an *in vitro* evolutionary process (e.g., SELEX (Systematic Evolution of Ligands by EXponential Enrichment), described in U.S. Patent No. 5,270,163), which selects for target-specific aptamer sequences from large combinatorial libraries. Aptamer compositions may be double-stranded or single-stranded, and may include deoxyribonucleotides, ribonucleotides, nucleotide derivatives, or other nucleotide-like molecules. The nucleotide components of an aptamer may have modified sugar groups (e.g., the 2'-OH group of a ribonucleotide may be replaced by 2'-F or 2'-NH<sub>2</sub>), which may improve a desired property, e.g., resistance to nucleases or longer lifetime in blood. Aptamers may be conjugated to other molecules, e.g., a high molecular weight carrier to slow clearance of the aptamer from the circulatory system. Aptamers may be specifically cross-linked to their cognate ligands, e.g., by photo-activation of a

cross-linker (Brody, E.N. and L. Gold (2000) J. Biotechnol. 74:5-13).

The term "intramer" refers to an aptamer which is expressed *in vivo*. For example, a vaccinia virus-based RNA expression system has been used to express specific RNA aptamers at high levels in the cytoplasm of leukocytes (Blind, M. et al. (1999) Proc. Natl. Acad. Sci. USA 96:3606-3610).

The term "spiegelmer" refers to an aptamer which includes L-DNA, L-RNA, or other left-handed nucleotide derivatives or nucleotide-like molecules. Aptamers containing left-handed nucleotides are resistant to degradation by naturally occurring enzymes, which normally act on substrates containing right-handed nucleotides.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a polynucleotide having a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic NAAP, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide" and a "composition comprising a given polypeptide" can refer to any composition containing the given polynucleotide or polypeptide. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotides encoding NAAP or fragments of NAAP may be employed as hybridization probes.



The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

5 "Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (Accelrys,  
10 Burlington MA) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino  
15 acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

	Original Residue	Conservative Substitution
	Ala	Gly, Ser
	Arg	His, Lys
20	Asn	Asp, Gln, His
	Asp	Asn, Glu
	Cys	Ala, Ser
	Gln	Asn, Glu, His
	Glu	Asp, Gln, His
25	Gly	Ala
	His	Asn, Arg, Gln, Glu
	Ile	Leu, Val
	Leu	Ile, Val
	Lys	Arg, Gln, Glu
30	Met	Leu, Ile
	Phe	His, Met, Leu, Trp, Tyr
	Ser	Cys, Thr
	Thr	Ser, Val
	Trp	Phe, Tyr
35	Tyr	His, Phe, Trp
	Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation,

(b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A “deletion” refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

5       The term “derivative” refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or  
10 immunological function of the polypeptide from which it was derived.

A “detectable label” refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

“Differential expression” refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such  
15 comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

“Exon shuffling” refers to the recombination of different coding regions (exons). Since an exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the  
20 evolution of new protein functions.

A “fragment” is a unique portion of NAAP or a polynucleotide encoding NAAP which can be identical in sequence to, but shorter in length than, the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from about 5 to about 1000 contiguous nucleotides or amino acid residues. A  
25 fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a  
30 certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:58-114 can comprise a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:58-114, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:58-114 can be employed in one or more embodiments of methods of the invention, for example, in hybridization and  
5 amplification technologies and in analogous methods that distinguish SEQ ID NO:58-114 from related polynucleotides. The precise length of a fragment of SEQ ID NO:58-114 and the region of SEQ ID NO:58-114 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-57 is encoded by a fragment of SEQ ID NO:58-114. A  
10 fragment of SEQ ID NO:1-57 can comprise a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-57. For example, a fragment of SEQ ID NO:1-57 can be used as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-57. The precise length of a fragment of SEQ ID NO:1-57 and the region of SEQ ID NO:1-57 to which the fragment corresponds can be determined based on the intended purpose for the fragment using  
15 one or more analytical methods described herein or otherwise known in the art.

A “full length” polynucleotide is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A “full length” polynucleotide sequence encodes a “full length” polypeptide sequence.

“Homology” refers to sequence similarity or, alternatively, sequence identity, between two or  
20 more polynucleotide sequences or two or more polypeptide sequences.

The terms “percent identity” and “% identity,” as applied to polynucleotide sequences, refer to the percentage of identical residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and  
25 therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using one or more computer algorithms or programs known in the art or described herein. For example, percent identity can be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the  
30 LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989; CABIOS 5:151-153) and in Higgins, D.G. et al. (1992; CABIOS 8:189-191). For pairwise alignments of

polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms which can be used is provided by the National Center for Biotechnology Information (NCBI) Basic  
 5 Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at <http://www.ncbi.nlm.nih.gov/BLAST/>. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2  
 10 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version  
 15 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

*Matrix: BLOSUM62*

*Reward for match: 1*

*Penalty for mismatch: -2*

*Open Gap: 5 and Extension Gap: 2 penalties*

20 *Gap x drop-off: 50*

*Expect: 10*

*Word Size: 11*

*Filter: on*

Percent identity may be measured over the length of an entire defined sequence, for example,  
 25 as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a  
 30 length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes

in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of identical residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide. The phrases "percent similarity" and "% similarity," as applied to polypeptide sequences, refer to the percentage of residue matches, including identical residue matches and conservative substitutions, between at least two polypeptide sequences aligned using a standardized algorithm. In contrast, conservative substitutions are not included in the calculation of percent identity between polypeptide sequences.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

*Matrix: BLOSUM62*  
*Open Gap: 11 and Extension Gap: 1 penalties*  
*Gap x drop-off: 50*  
*Expect: 10*  
*Word Size: 3*  
*Filter: on*

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for

instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

5       “Human artificial chromosomes” (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

      The term “humanized antibody” refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely  
10 resembles a human antibody, and still retains its original binding ability.

      “Hybridization” refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized  
15 after the “washing” step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be  
20 varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

      Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about  
25 5°C to 20°C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic strength and pH. The  $T_m$  is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating  $T_m$  and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. and D.W. Russell (2001; Molecular Cloning: A Laboratory Manual, 3rd ed., vol. 1-3, Cold Spring Harbor Press,  
30 Cold Spring Harbor NY, ch. 9).

      High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour.

Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

10       The term “hybridization complex” refers to a complex formed between two nucleic acids by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C<sub>0</sub>t or R<sub>0</sub>t analysis) or formed between one nucleic acid present in solution and another nucleic acid immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been  
15       fixed).

      The words “insertion” and “addition” refer to changes in an amino acid or polynucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

      “Immune response” can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression  
20       of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

      An “immunogenic fragment” is a polypeptide or oligopeptide fragment of NAAP which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term “immunogenic fragment” also includes any polypeptide or oligopeptide fragment  
25       of NAAP which is useful in any of the antibody production methods disclosed herein or known in the art.

      The term “microarray” refers to an arrangement of a plurality of polynucleotides, polypeptides, antibodies, or other chemical compounds on a substrate.

      The terms “element” and “array element” refer to a polynucleotide, polypeptide, antibody, or  
30       other chemical compound having a unique and defined position on a microarray.

      The term “modulate” refers to a change in the activity of NAAP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological,

functional, or immunological properties of NAAP.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an NAAP may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of NAAP.

"Probe" refers to nucleic acids encoding NAAP, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acids. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the



specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in, for example, Sambrook, J. and D.W. Russell (2001; Molecular Cloning: A Laboratory Manual, 3rd ed., vol. 1-3, Cold Spring Harbor Press, Cold Spring Harbor NY), Ausubel, F.M. et al. (1999; Short Protocols in Molecular Biology, 4<sup>th</sup> ed., John Wiley & Sons, New York NY), and Innis, M. et al. (1990; PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA). PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a nucleic acid that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of

sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook and Russell (*supra*). The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the  
5 nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be used to vaccinate a mammal wherein the recombinant nucleic acid is  
10 expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

15 "Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA molecule, is composed of the same linear  
20 sequence of nucleotides as the reference DNA molecule with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing NAAP, nucleic acids encoding NAAP, or fragments thereof may comprise a bodily fluid; an extract from a  
25 cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure  
30 of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the

antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least about 60% free, preferably at least about 75% free, and most preferably at least about 90% free from other  
5 components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers,  
10 microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" or "expression profile" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient  
15 cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed  
20 cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid  
25 introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. In another embodiment, the nucleic acid can be introduced by infection with a recombinant viral vector, such as a lentiviral vector (Lois, C. et al. (2002) Science 295:868-872). The  
30 term genetic manipulation does not include classical cross-breeding, or *in vitro* fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants

and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook and Russell (*supra*).

5       A “variant” of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the “BLAST 2 Sequences” tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least  
10 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an “allelic” (as defined above), “splice,” “species,” or “polymorphic” variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding  
15 polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotides that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass “single nucleotide  
20 polymorphisms” (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

      A “variant” of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity or sequence similarity to the particular polypeptide sequence over a  
25 certain length of one of the polypeptide sequences using blastp with the “BLAST 2 Sequences” tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity or sequence similarity over a certain defined length of one  
30 of the polypeptides.

## THE INVENTION

Various embodiments of the invention include new human nucleic acid-associated proteins (NAAP), the polynucleotides encoding NAAP, and the use of these compositions for the diagnosis, treatment, or prevention of cell proliferative, neurological, developmental, and autoimmune/inflammatory disorders, and infections.

5           Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide embodiments of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide  
10       sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown. Column 6 shows the Incyte ID numbers of physical, full length clones corresponding to the polypeptide and polynucleotide sequences of the invention. The full length clones encode polypeptides which have at least 95% sequence identity to the polypeptide sequences shown in column 3.

15           Table 2 shows sequences with homology to polypeptide embodiments of the invention as identified by BLAST analysis against the GenBank protein (genpept) database and the PROTEOME database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (GenBank ID NO:) of the nearest GenBank homolog and the PROTEOME database identification numbers  
20       (PROTEOME ID NO:) of the nearest PROTEOME database homologs. Column 4 shows the probability scores for the matches between each polypeptide and its homolog(s). Column 5 shows the annotation of the GenBank and PROTEOME database homolog(s) along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

25           Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS  
30       program of the GCG sequence analysis software package (Accelrys, Burlington MA). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to

which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are nucleic acid-associated proteins. For example, SEQ ID NO:11 is 100% identical, from residue M1 to residue L174 and from residue Q175 to residue Q488, to a human protein similar to interferon regulatory factor 5 (GenBank ID g13278720) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is  $5.8e-273$ , which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:11 also has homology to human and mouse interferon regulatory factor 5, a transcription factor that induces expression of members of the interferon A family in response to NDV viral infection, as determined by BLAST analysis using the PROTEOME database. SEQ ID NO:11 also contains an interferon regulatory factor transcription factor domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:11 is a member of the interferon regulatory factor (IRF) family of transcription factors.

In an alternative example, SEQ ID NO:24 is 96% identical, from residue M1 to residue R133, to human hMBF1alpha (GenBank ID g6526355) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is  $5.0e-63$ , which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:24 also has homology to endothelial differentiation-related factor-1, a putative transcriptional coactivator that binds to calmodulin (CALM1), in a calcium-dependent manner, and to the TATA-binding protein (TBP), has transcription factor function, and are EDF1 transcription coactivators, as determined by BLAST analysis using the PROTEOME database. SEQ ID NO:24 also contains a helix-turn-helix domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from additional BLAST analyses provide further corroborative evidence that SEQ ID NO:24 is a transcription factor regulator.

In an alternative example, SEQ ID NO:35 is 84% identical, from residue M932 to residue G1638, to human ras-responsive element binding protein (GenBank ID g1654112) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:35 also has homology to nuclear transcription factors, as determined by BLAST

analysis using the PROTEOME database. SEQ ID NO:35 also contains a zinc finger domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and MOTIFS analyses provide further corroborative evidence that SEQ ID NO:35 is a transcription  
 5 factor.

In an alternative example, SEQ ID NO:55 is 85% identical, from residue M1 to residue V508, to rat serine-arginine-rich splicing regulatory protein, SRRP86 (GenBank ID g7158880) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is  $4.8e-225$ , which indicates the probability of obtaining the observed polypeptide sequence alignment  
 10 by chance. SEQ ID NO:55 also has homology to proteins that are localized to the nucleus, function as RNA-binding proteins, and are serine-arginine-rich proteins, as determined by BLAST analysis using the PROTEOME database. SEQ ID NO:55 also contains an RNA recognition motif as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, and other  
 15 BLAST analyses provide further corroborative evidence that SEQ ID NO:55 is a serine-arginine-rich splicing regulatory protein.

SEQ ID NO:1-10, SEQ ID NO:12-23, SEQ ID NO:25-34, SEQ ID NO:36-54 and SEQ ID NO:56-57 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-57 are described in Table 7.

20 As shown in Table 4, the full length polynucleotide embodiments were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Column 1 lists the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:), the corresponding Incyte polynucleotide consensus sequence number (Incyte ID) for each polynucleotide of the invention, and the length of each polynucleotide sequence in basepairs.  
 25 Column 2 shows the nucleotide start (5') and stop (3') positions of the cDNA and/or genomic sequences used to assemble the full length polynucleotide embodiments, and of fragments of the polynucleotides which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:58-114 or that distinguish between SEQ ID NO:58-114 and related polynucleotides.

30 The polynucleotide fragments described in Column 2 of Table 4 may refer specifically, for example, to Incyte cDNAs derived from tissue-specific cDNA libraries or from pooled cDNA libraries. Alternatively, the polynucleotide fragments described in column 2 may refer to GenBank

cDNAs or ESTs which contributed to the assembly of the full length polynucleotides. In addition, the polynucleotide fragments described in column 2 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the polynucleotide fragments described in column 2 may be derived from the

5 NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation "NP"). Alternatively, the polynucleotide fragments described in column 2 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, a polynucleotide sequence identified as

10 FL\_XXXXXX\_N<sub>1</sub>\_N<sub>2</sub>\_YYYYY\_N<sub>3</sub>\_N<sub>4</sub> represents a "stitched" sequence in which XXXXXX is the identification number of the cluster of sequences to which the algorithm was applied, and YYYYY is the number of the prediction generated by the algorithm, and N<sub>1,2,3...</sub>, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the polynucleotide fragments in column 2 may refer to assemblages of exons brought together by an

15 "exon-stretching" algorithm. For example, a polynucleotide sequence identified as FLXXXXXX\_gAAAAA\_gBBBBB\_1\_N is a "stretched" sequence, with XXXXXX being the Incyte project identification number, gAAAAA being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, gBBBBB being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog,

20 and N referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the "exon-stretching" algorithm, a RefSeq identifier (denoted by "NM," "NP," or "NT") may be used in place of the GenBank identifier (*i.e.*, gBBBBB).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The

25 following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

Prefix	Type of analysis and/or examples of programs
GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK).
GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).



INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.
------	---

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in Table 4 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotides which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotides. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

Table 8 shows single nucleotide polymorphisms (SNPs) found in polynucleotide sequences of the invention, along with allele frequencies in different human populations. Columns 1 and 2 show the polynucleotide sequence identification number (SEQ ID NO:) and the corresponding Incyte project identification number (PID) for polynucleotides of the invention. Column 3 shows the Incyte identification number for the EST in which the SNP was detected (EST ID), and column 4 shows the identification number for the SNP (SNP ID). Column 5 shows the position within the EST sequence at which the SNP is located (EST SNP), and column 6 shows the position of the SNP within the full-length polynucleotide sequence (CB1 SNP). Column 7 shows the allele found in the EST sequence. Columns 8 and 9 show the two alleles found at the SNP site. Column 10 shows the amino acid encoded by the codon including the SNP site, based upon the allele found in the EST. Columns 11-14 show the frequency of allele 1 in four different human populations. An entry of n/d (not detected) indicates that the frequency of allele 1 in the population was too low to be detected, while n/a (not available) indicates that the allele frequency was not determined for the population.

The invention also encompasses NAAP variants. Various embodiments of NAAP variants can have at least about 80%, at least about 90%, or at least about 95% amino acid sequence identity to the NAAP amino acid sequence, and can contain at least one functional or structural characteristic of NAAP.

Various embodiments also encompass polynucleotides which encode NAAP. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:58-114, which encodes NAAP. The polynucleotide

sequences of SEQ ID NO:58-114, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses variants of a polynucleotide encoding NAAP. In particular, such a variant polynucleotide will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a polynucleotide encoding NAAP. A particular aspect of the invention encompasses a variant of a polynucleotide comprising a sequence selected from the group consisting of SEQ ID NO:58-114 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:58-114. Any one of the polynucleotide variants described above can encode a polypeptide which contains at least one functional or structural characteristic of NAAP.

In addition, or in the alternative, a polynucleotide variant of the invention is a splice variant of a polynucleotide encoding NAAP. A splice variant may have portions which have significant sequence identity to a polynucleotide encoding NAAP, but will generally have a greater or lesser number of polynucleotides due to additions or deletions of blocks of sequence arising from alternate splicing of exons during mRNA processing. A splice variant may have less than about 70%, or alternatively less than about 60%, or alternatively less than about 50% polynucleotide sequence identity to a polynucleotide encoding NAAP over its entire length; however, portions of the splice variant will have at least about 70%, or alternatively at least about 85%, or alternatively at least about 95%, or alternatively 100% polynucleotide sequence identity to portions of the polynucleotide encoding NAAP. Any one of the splice variants described above can encode a polypeptide which contains at least one functional or structural characteristic of NAAP.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding NAAP, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring NAAP, and all such variations are to be considered as being specifically disclosed.

Although polynucleotides which encode NAAP and its variants are generally capable of

hybridizing to polynucleotides encoding naturally occurring NAAP under appropriately selected conditions of stringency, it may be advantageous to produce polynucleotides encoding NAAP or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding NAAP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

10       The invention also encompasses production of polynucleotides which encode NAAP and NAAP derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic polynucleotide may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a polynucleotide encoding NAAP or any fragment thereof.

15       Embodiments of the invention can also include polynucleotides that are capable of hybridizing to the claimed polynucleotides, and, in particular, to those having the sequences shown in SEQ ID NO:58-114 and fragments thereof, under various conditions of stringency (Wahl, G.M. and S.L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-511). Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

20       Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Biosciences, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Invitrogen, Carlsbad CA). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Amersham Biosciences), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art (Ausubel et al., *supra*, ch. 7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853).

25  
30

The nucleic acids encoding NAAP may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector (Sarkar, G. (1993) PCR Methods Applic. 2:318-322). Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences (Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186). A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA (Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119). In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art (Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer

controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotides or fragments thereof which encode NAAP may be cloned in recombinant DNA molecules that direct expression of NAAP, or fragments  
5 or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other polynucleotides which encode substantially the same or a functionally equivalent polypeptides may be produced and used to express NAAP.

The polynucleotides of the invention can be engineered using methods generally known in the art in order to alter NAAP-encoding sequences for a variety of purposes including, but not limited to,  
10 modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

15 The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent No. 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Cramer, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of NAAP, such as its biological or enzymatic activity or its ability to bind to  
20 other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular  
25 evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

30 In another embodiment, polynucleotides encoding NAAP may be synthesized, in whole or in part, using one or more chemical methods well known in the art (Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232).

Alternatively, NAAP itself or a fragment thereof may be synthesized using chemical methods known in the art. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques (Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; Roberge, J.Y. et al. (1995) *Science* 269:202-204). Automated  
5 synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of NAAP, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid  
10 chromatography (Chiez, R.M. and F.Z. Regnier (1990) *Methods Enzymol.* 182:392-421). The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing (Creighton, *supra*, pp. 28-53).

In order to express a biologically active NAAP, the polynucleotides encoding NAAP or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains  
15 the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotides encoding NAAP. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of polynucleotides encoding NAAP. Such signals include  
20 the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where a polynucleotide sequence encoding NAAP and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be  
25 provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used (Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162).

Methods which are well known to those skilled in the art may be used to construct expression  
30 vectors containing polynucleotides encoding NAAP and appropriate transcriptional and translational control elements. These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination (Sambrook and Russell, *supra*, ch. 1-4, and 8; Ausubel et al.,

*supra*, ch. 1, 3, and 15).

A variety of expression vector/host systems may be utilized to contain and express polynucleotides encoding NAAP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems (Sambrook and Russell, *supra*; Ausubel et al., *supra*; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355). Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of polynucleotides to the targeted organ, tissue, or cell population (Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5:350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90:6340-6344; Buller, R.M. et al. (1985) Nature 317:813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31:219-226; Verma, I.M. and N. Somia (1997) Nature 389:239-242). The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotides encoding NAAP. For example, routine cloning, subcloning, and propagation of polynucleotides encoding NAAP can be achieved using a multifunctional *E. coli* vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPO1 plasmid (Invitrogen). Ligation of polynucleotides encoding NAAP into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for *in vitro* transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence (Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509). When large quantities of NAAP are needed, e.g. for the production of antibodies, vectors which direct high level expression of NAAP may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of NAAP. A number of vectors

containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast *Saccharomyces cerevisiae* or *Pichia pastoris*. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign polynucleotide sequences into the host genome for stable propagation (Ausubel et al.,  
5 *supra*; Bitter, G.A. et al. (1987) *Methods Enzymol.* 153:516-544; Scorer, C.A. et al. (1994) *Bio/Technology* 12:181-184).

Plant systems may also be used for expression of NAAP. Transcription of polynucleotides encoding NAAP may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) *EMBO J.*  
10 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used (Coruzzi, G. et al. (1984) *EMBO J.* 3:1671-1680; Broglie, R. et al. (1984) *Science* 224:838-843; Winter, J. et al. (1991) *Results Probl. Cell Differ.* 17:85-105). These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection (The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp.  
15 191-196).

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, polynucleotides encoding NAAP may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain  
20 infective virus which expresses NAAP in host cells (Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:3655-3659). In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of  
25 DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes (Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355).

For long term production of recombinant proteins in mammalian systems, stable expression of NAAP in cell lines is preferred. For example, polynucleotides encoding NAAP can be transformed  
30 into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched



media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

5 Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk* and *apr* cells, respectively (Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823). Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to  
10 methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14). Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites (Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA  
15 85:8047-8051). Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech),  $\beta$ -glucuronidase and its substrate  $\beta$ -glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131).

20 Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding NAAP is inserted within a marker gene sequence, transformed cells containing polynucleotides encoding NAAP can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding NAAP under the  
25 control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the polynucleotide encoding NAAP and that express NAAP may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and  
30 protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of NAAP using either

specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on NAAP is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art (Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ).

10 A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding NAAP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, polynucleotides encoding NAAP, or any fragments thereof, may be cloned into a vector 15 for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Biosciences, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of 20 detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with polynucleotides encoding NAAP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence 25 and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode NAAP may be designed to contain signal sequences which direct secretion of NAAP through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted polynucleotides or to process the expressed protein in the desired fashion. Such modifications 30 of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity.

Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

5 In another embodiment of the invention, natural, modified, or recombinant polynucleotides encoding NAAP may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric NAAP protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of NAAP activity. Heterologous protein and peptide  
10 moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins,  
15 respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the NAAP encoding sequence and the heterologous protein sequence, so that NAAP may be cleaved away from the heterologous moiety following purification. Methods for fusion protein  
20 expression and purification are discussed in Ausubel et al. (*supra*, ch. 10 and 16). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In another embodiment, synthesis of radiolabeled NAAP may be achieved *in vitro* using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6  
25 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, <sup>35</sup>S-methionine.

NAAP, fragments of NAAP, or variants of NAAP may be used to screen for compounds that specifically bind to NAAP. One or more test compounds may be screened for specific binding to NAAP. In various embodiments, 1, 2, 3, 4, 5, 10, 20, 50, 100, or 200 test compounds can be screened  
30 for specific binding to NAAP. Examples of test compounds can include antibodies, anticalins, oligonucleotides, proteins (e.g., ligands or receptors), or small molecules.

In related embodiments, variants of NAAP can be used to screen for binding of test

compounds, such as antibodies, to NAAP, a variant of NAAP, or a combination of NAAP and/or one or more variants NAAP. In an embodiment, a variant of NAAP can be used to screen for compounds that bind to a variant of NAAP, but not to NAAP having the exact sequence of a sequence of SEQ ID NO:1-57. NAAP variants used to perform such screening can have a range of  
5 about 50% to about 99% sequence identity to NAAP, with various embodiments having 60%, 70%, 75%, 80%, 85%, 90%, and 95% sequence identity.

In an embodiment, a compound identified in a screen for specific binding to NAAP can be closely related to the natural ligand of NAAP, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner (Coligan, J.E. et al. (1991) Current  
10 Protocols in Immunology 1(2):Chapter 5). In another embodiment, the compound thus identified can be a natural ligand of a receptor NAAP (Howard, A.D. et al. (2001) Trends Pharmacol. Sci.22:132-140; Wise, A. et al. (2002) Drug Discovery Today 7:235-246).

In other embodiments, a compound identified in a screen for specific binding to NAAP can be closely related to the natural receptor to which NAAP binds, at least a fragment of the receptor, or a  
15 fragment of the receptor including all or a portion of the ligand binding site or binding pocket. For example, the compound may be a receptor for NAAP which is capable of propagating a signal, or a decoy receptor for NAAP which is not capable of propagating a signal (Ashkenazi, A. and V.M. Divit (1999) Curr. Opin. Cell Biol. 11:255-260; Mantovani, A. et al. (2001) Trends Immunol. 22:328-336). The compound can be rationally designed using known techniques. Examples of such techniques  
20 include those used to construct the compound etanercept (ENBREL; Amgen Inc., Thousand Oaks CA), which is efficacious for treating rheumatoid arthritis in humans. Etanercept is an engineered p75 tumor necrosis factor (TNF) receptor dimer linked to the Fc portion of human IgG<sub>1</sub> (Taylor, P.C. et al. (2001) Curr. Opin. Immunol. 13:611-616).

In one embodiment, two or more antibodies having similar or, alternatively, different  
25 specificities can be screened for specific binding to NAAP, fragments of NAAP, or variants of NAAP. The binding specificity of the antibodies thus screened can thereby be selected to identify particular fragments or variants of NAAP. In one embodiment, an antibody can be selected such that its binding specificity allows for preferential identification of specific fragments or variants of NAAP. In another embodiment, an antibody can be selected such that its binding specificity allows for  
30 preferential diagnosis of a specific disease or condition having increased, decreased, or otherwise abnormal production of NAAP.

In an embodiment, anticalins can be screened for specific binding to NAAP, fragments of

NAAP, or variants of NAAP. Anticalins are ligand-binding proteins that have been constructed based on a lipocalin scaffold (Weiss, G.A. and H.B. Lowman (2000) Chem. Biol. 7:R177-R184; Skerra, A. (2001) J. Biotechnol. 74:257-275). The protein architecture of lipocalins can include a beta-barrel having eight antiparallel beta-strands, which supports four loops at its open end. These loops form the natural ligand-binding site of the lipocalins, a site which can be re-engineered *in vitro* by amino acid substitutions to impart novel binding specificities. The amino acid substitutions can be made using methods known in the art or described herein, and can include conservative substitutions (e.g., substitutions that do not alter binding specificity) or substitutions that modestly, moderately, or significantly alter binding specificity.

10 In one embodiment, screening for compounds which specifically bind to, stimulate, or inhibit NAAP involves producing appropriate cells which express NAAP, either as a secreted protein or on the cell membrane. Preferred cells can include cells from mammals, yeast, *Drosophila*, or *E. coli*. Cells expressing NAAP or cell membrane fractions which contain NAAP are then contacted with a test compound and binding, stimulation, or inhibition of activity of either NAAP or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with NAAP, either in solution or affixed to a solid support, and detecting the binding of NAAP to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

25 An assay can be used to assess the ability of a compound to bind to its natural ligand and/or to inhibit the binding of its natural ligand to its natural receptors. Examples of such assays include radio-labeling assays such as those described in U.S. Patent No. 5,914,236 and U.S. Patent No. 6,372,724. In a related embodiment, one or more amino acid substitutions can be introduced into a polypeptide compound (such as a receptor) to improve or alter its ability to bind to its natural ligands (Matthews, D.J. and J.A. Wells. (1994) Chem. Biol. 1:25-30). In another related embodiment, one or more amino acid substitutions can be introduced into a polypeptide compound (such as a ligand) to improve or alter its ability to bind to its natural receptors (Cunningham, B.C. and J.A. Wells (1991) Proc. Natl. Acad. Sci. USA 88:3407-3411; Lowman, H.B. et al. (1991) J. Biol. Chem. 266:10982-10988).

30 NAAP, fragments of NAAP, or variants of NAAP may be used to screen for compounds

that modulate the activity of NAAP. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for NAAP activity, wherein NAAP is combined with at least one test compound, and the activity of NAAP in the presence of a test compound is compared with the activity of NAAP in the absence of the test compound. A change in the activity of NAAP in the presence of the test compound is indicative of a compound that modulates the activity of NAAP. Alternatively, a test compound is combined with an *in vitro* or cell-free system comprising NAAP under conditions suitable for NAAP activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of NAAP may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding NAAP or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease (see, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337). For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (*neo*; Capecchi, M.R. (1989) *Science* 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) *Clin. Invest.* 97:1999-2002; Wagner, K.U. et al. (1997) *Nucleic Acids Res.* 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding NAAP may also be manipulated *in vitro* in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) *Science* 282:1145-1147).

Polynucleotides encoding NAAP can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region

of a polynucleotide encoding NAAP is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, 5 a mammal inbred to overexpress NAAP, e.g., by secreting NAAP in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) *Biotechnol. Annu. Rev.* 4:55-74).

## THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of NAAP and nucleic acid-associated proteins. In addition, examples of tissues expressing 10 NAAP can be found in Table 6 and can also be found in Example XI. Therefore, NAAP appears to play a role in cell proliferative, neurological, developmental, and autoimmune/inflammatory disorders, and infections. In the treatment of disorders associated with increased NAAP expression or activity, it is desirable to decrease the expression or activity of NAAP. In the treatment of disorders associated with decreased NAAP expression or activity, it is desirable to increase the expression or 15 activity of NAAP.

Therefore, in one embodiment, NAAP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of NAAP. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed 20 connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, a cancer of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, colon, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, 25 skin, spleen, testis, thymus, thyroid, and uterus; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral 30 meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial

insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorder of the central nervous system, cerebral palsy, a neuroskeletal disorder, an autonomic nervous system disorder, a cranial nerve disorder, a spinal cord disease, muscular dystrophy and other neuromuscular disorder, a peripheral nervous system disorder, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathy, myasthenia gravis, periodic paralysis, a mental disorder including mood, anxiety, and schizophrenic disorder, seasonal affective disorder (SAD), akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, and Tourette's disorder; a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; and an infection, such as those caused by a viral agent classified as adenovirus, arenavirus, bunyavirus, calicivirus, coronavirus, filovirus, hepadnavirus, herpesvirus, flavivirus, orthomyxovirus, parvovirus, papovavirus, paramyxovirus, picornavirus, poxvirus, reovirus, retrovirus, rhabdovirus, or togavirus; an infection



caused by a bacterial agent classified as pneumococcus, staphylococcus, streptococcus, bacillus, corynebacterium, clostridium, meningococcus, gonococcus, listeria, moraxella, kingella, haemophilus, legionella, bordetella, gram-negative enterobacterium including shigella, salmonella, or campylobacter, pseudomonas, vibrio, brucella, francisella, yersinia, bartonella, norcardium, actinomyces, mycobacterium, spirochaetale, rickettsia, chlamydia, or mycoplasma; an infection caused by a fungal agent classified as aspergillus, blastomyces, dermatophytes, cryptococcus, coccidioides, malassezia, histoplasma, or other mycosis-causing fungal agent; and an infection caused by a parasite classified as plasmodium or malaria-causing, parasitic entamoeba, leishmania, trypanosoma, toxoplasma, pneumocystis carinii, intestinal protozoa such as giardia, trichomonas, tissue nematode such as trichinella, intestinal nematode such as ascaris, lymphatic filarial nematode, trematode such as schistosoma, and cestode such as tapeworm.

In another embodiment, a vector capable of expressing NAAP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of NAAP including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified NAAP in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of NAAP including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of NAAP may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of NAAP including, but not limited to, those listed above.

In a further embodiment, an antagonist of NAAP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of NAAP. Examples of such disorders include, but are not limited to, those cell proliferative, neurological, developmental, and autoimmune/inflammatory disorders, and infections described above. In one aspect, an antibody which specifically binds NAAP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express NAAP.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding NAAP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of NAAP including, but not limited to, those described above.

In other embodiments, any protein, agonist, antagonist, antibody, complementary sequence, or vector embodiments may be administered in combination with other appropriate therapeutic agents.

Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of NAAP may be produced using methods which are generally known in the art. In particular, purified NAAP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind NAAP. Antibodies to NAAP may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. In an embodiment, neutralizing antibodies (i.e., those which inhibit dimer formation) can be used therapeutically. Single chain antibodies (e.g., from camels or llamas) may be potent enzyme inhibitors and may have application in the design of peptide mimetics, and in the development of immuno-adsorbents and biosensors (Muyldermans, S. (2001) *J. Biotechnol.* 74:277-302).

For the production of antibodies, various hosts including goats, rabbits, rats, mice, camels, dromedaries, llamas, humans, and others may be immunized by injection with NAAP or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and *Corynebacterium parvum* are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to NAAP have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are substantially identical to a portion of the amino acid sequence of the natural protein. Short stretches of NAAP amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to NAAP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma

technique (Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120).

In addition, techniques developed for the production of "chimeric antibodies," such as the  
5 splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used (Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; Takeda, S. et al. (1985) Nature 314:452-454). Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce NAAP-specific single chain antibodies.  
10 Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries (Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137).

Antibodies may also be produced by inducing *in vivo* production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in  
15 the literature (Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299).

Antibody fragments which contain specific binding sites for NAAP may also be generated. For example, such fragments include, but are not limited to, F(ab')<sub>2</sub> fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of  
20 the F(ab')<sub>2</sub> fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity (Huse, W.D. et al. (1989) Science 246:1275-1281).

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either  
25 polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between NAAP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering NAAP epitopes is generally used, but a competitive binding assay may also be employed (Pound, *supra*).

30 Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for NAAP. Affinity is expressed as an association constant,  $K_a$ , which is defined as the molar concentration of NAAP-antibody complex divided by the

molar concentrations of free antigen and free antibody under equilibrium conditions. The  $K_a$  determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple NAAP epitopes, represents the average affinity, or avidity, of the antibodies for NAAP. The  $K_a$  determined for a preparation of monoclonal antibodies, which are monospecific for a particular NAAP epitope, represents a true measure of affinity. High-affinity antibody preparations with  $K_a$  ranging from about  $10^9$  to  $10^{12}$  L/mole are preferred for use in immunoassays in which the NAAP-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with  $K_a$  ranging from about  $10^6$  to  $10^7$  L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of NAAP, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of NAAP-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available (Catty, *supra*; Coligan et al., *supra*).

In another embodiment of the invention, polynucleotides encoding NAAP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding NAAP. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding NAAP (Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press, Totawa NJ).

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein (Slater, J.E. et al. (1998) *J. Allergy Clin. Immunol.* 102:469-475; Scanlon, K.J. et al. (1995) 9:1288-1296). Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and

adeno-associated virus vectors (Miller, A.D. (1990) *Blood* 76:271; Ausubel et al., *supra*; Uckert, W. and W. Walther (1994) *Pharmacol. Ther.* 63:323-347). Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art (Rossi, J.J. (1995) *Br. Med. Bull.* 51:217-225; Boado, R.J. et al. (1998) *J. Pharm. Sci.* 87:1308-1315; Morris, M.C. et al. (1997) *Nucleic Acids Res.* 25:2730-2736).

In another embodiment of the invention, polynucleotides encoding NAAP may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) *Science* 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaise, R.M. et al. (1995) *Science* 270:475-480; Bordignon, C. et al. (1995) *Science* 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) *Cell* 75:207-216; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:643-666; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) *Science* 270:404-410; Verma, I.M. and N. Somia (1997) *Nature* 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) *Nature* 335:395-396; Poeschla, E. et al. (1996) *Proc. Natl. Acad. Sci. USA* 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as *Candida albicans* and *Paracoccidioides brasiliensis*; and protozoan parasites such as *Plasmodium falciparum* and *Trypanosoma cruzi*). In the case where a genetic deficiency in NAAP expression or regulation causes disease, the expression of NAAP from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in NAAP are treated by constructing mammalian expression vectors encoding NAAP and introducing these vectors by mechanical means into NAAP-deficient cells. Mechanical transfer technologies for use with cells *in vivo* or *ex vitro* include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) *Annu. Rev. Biochem.* 62:191-217; Ivics, Z. (1997) *Cell* 91:501-510; Boulay, J.-L. and H. Récipon (1998) *Curr. Opin. Biotechnol.* 9:445-450).

Expression vectors that may be effective for the expression of NAAP include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). NAAP  
 5 may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or  $\beta$ -actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen));  
 10 the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and H.M. Blau, *supra*), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding NAAP from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID  
 15 TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these  
 20 standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to NAAP expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding NAAP under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive  
 25 element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for  
 30 receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et

al. (1998) J. Virol. 72:9873-9880). U.S. Patent No. 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4<sup>+</sup> T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In an embodiment, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding NAAP to cells which have one or more genetic abnormalities with respect to the expression of NAAP. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent No. 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999; Annu. Rev. Nutr. 19:511-544) and Verma, I.M. and N. Somia (1997; Nature 18:389:239-242).

In another embodiment, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding NAAP to target cells which have one or more genetic abnormalities with respect to the expression of NAAP. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing NAAP to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent No. 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent No. 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999; J. Virol. 73:519-532) and Xu, H. et al. (1994; Dev. Biol. 163:152-161). The manipulation of cloned herpesvirus sequences, the generation of

recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another embodiment, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding NAAP to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) *Curr. Opin. Biotechnol.* 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for NAAP into the alphavirus genome in place of the capsid-coding region results in the production of a large number of NAAP-coding RNAs and the synthesis of high levels of NAAP in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) *Virology* 228:74-83). The wide host range of alphaviruses will allow the introduction of NAAP into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature (Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177). A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of



RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of RNA molecules encoding NAAP.

5        Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of  
10 candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively,  
15 RNA molecules may be generated by *in vitro* and *in vivo* transcription of DNA molecules encoding NAAP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible  
20 modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine,  
25 guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

In other embodiments of the invention, the expression of one or more selected polynucleotides of the present invention can be altered, inhibited, decreased, or silenced using RNA interference (RNAi) or post-transcriptional gene silencing (PTGS) methods known in the art. RNAi is a post-transcriptional mode of gene silencing in which double-stranded RNA (dsRNA) introduced into a  
30 targeted cell specifically suppresses the expression of the homologous gene (i.e., the gene bearing the sequence complementary to the dsRNA). This effectively knocks out or substantially reduces the expression of the targeted gene. PTGS can also be accomplished by use of DNA or DNA fragments

as well. RNAi methods are described by Fire, A. et al. (1998; Nature 391:806-811) and Gura, T. (2000; Nature 404:804-808). PTGS can also be initiated by introduction of a complementary segment of DNA into the selected tissue using gene delivery and/or viral vector delivery methods described herein or known in the art.

5 RNAi can be induced in mammalian cells by the use of small interfering RNA also known as siRNA. SiRNA are shorter segments of dsRNA (typically about 21 to 23 nucleotides in length) that result *in vivo* from cleavage of introduced dsRNA by the action of an endogenous ribonuclease. SiRNA appear to be the mediators of the RNAi effect in mammals. The most effective siRNAs appear to be 21 nucleotide dsRNAs with 2 nucleotide 3' overhangs. The use of siRNA for inducing  
10 RNAi in mammalian cells is described by Elbashir, S.M. et al. (2001; Nature 411:494-498).

SiRNA can either be generated indirectly by introduction of dsRNA into the targeted cell, or directly by mammalian transfection methods and agents described herein or known in the art (such as liposome-mediated transfection, viral vector methods, or other polynucleotide delivery/introductory methods). Suitable SiRNAs can be selected by examining a transcript of the target polynucleotide  
15 (e.g., mRNA) for nucleotide sequences downstream from the AUG start codon and recording the occurrence of each nucleotide and the 3' adjacent 19 to 23 nucleotides as potential siRNA target sites, with sequences having a 21 nucleotide length being preferred. Regions to be avoided for target siRNA sites include the 5' and 3' untranslated regions (UTRs) and regions near the start codon (within 75 bases), as these may be richer in regulatory protein binding sites. UTR-binding proteins and/or  
20 translation initiation complexes may interfere with binding of the siRNP endonuclease complex. The selected target sites for siRNA can then be compared to the appropriate genome database (e.g., human, etc.) using BLAST or other sequence comparison algorithms known in the art. Target sequences with significant homology to other coding sequences can be eliminated from consideration. The selected SiRNAs can be produced by chemical synthesis methods known in the art or by *in vitro*  
25 transcription using commercially available methods and kits such as the SILENCER siRNA construction kit (Ambion, Austin TX).

In alternative embodiments, long-term gene silencing and/or RNAi effects can be induced in selected tissue using expression vectors that continuously express siRNA. This can be accomplished using expression vectors that are engineered to express hairpin RNAs (shRNAs) using methods  
30 known in the art (see, e.g., Brummelkamp, T.R. et al. (2002) Science 296:550-553; and Paddison, P.J. et al. (2002) Genes Dev. 16:948-958). In these and related embodiments, shRNAs can be delivered to target cells using expression vectors known in the art. An example of a suitable expression vector for

delivery of siRNA is the PSILENCER1.0-U6 (circular) plasmid (Ambion). Once delivered to the target tissue, shRNAs are processed *in vivo* into siRNA-like molecules capable of carrying out gene-specific silencing.

In various embodiments, the expression levels of genes targeted by RNAi or PTGS methods  
5 can be determined by assays for mRNA and/or protein analysis. Expression levels of the mRNA of a targeted gene, can be determined by northern analysis methods using, for example, the NORTHERNMAX-GLY kit (Ambion); by microarray methods; by PCR methods; by real time PCR methods; and by other RNA/polynucleotide assays known in the art or described herein. Expression levels of the protein encoded by the targeted gene can be determined by Western analysis using  
10 standard techniques known in the art.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding NAAP. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides,  
15 transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased NAAP expression or activity, a compound which specifically inhibits expression of the polynucleotide  
20 encoding NAAP may be therapeutically useful, and in the treatment of disorders associated with decreased NAAP expression or activity, a compound which specifically promotes expression of the polynucleotide encoding NAAP may be therapeutically useful.

In various embodiments, one or more test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method  
25 commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a  
30 polynucleotide encoding NAAP is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an *in vitro* cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding NAAP are assayed by

any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding NAAP. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a *Schizosaccharomyces pombe* gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruce, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruce, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use *in vivo*, *in vitro*, and *ex vivo*. For *ex vivo* therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art (Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466).

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of NAAP, antibodies to NAAP, and mimetics, agonists, antagonists, or inhibitors of NAAP.

In various embodiments, the compositions described herein, such as pharmaceutical compositions, may be administered by any number of routes including, but not limited to, oral,

intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery allows administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising NAAP or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, NAAP or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example NAAP or fragments thereof, antibodies of NAAP, and agonists, antagonists or inhibitors of NAAP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED<sub>50</sub> (the dose therapeutically effective in 50% of the population) or LD<sub>50</sub> (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD<sub>50</sub>/ED<sub>50</sub> ratio. Compositions which exhibit large

therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED<sub>50</sub> with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1  $\mu$ g to 100,000  $\mu$ g, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

## DIAGNOSTICS

In another embodiment, antibodies which specifically bind NAAP may be used for the diagnosis of disorders characterized by expression of NAAP, or in assays to monitor patients being treated with NAAP or agonists, antagonists, or inhibitors of NAAP. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for NAAP include methods which utilize the antibody and a label to detect NAAP in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring NAAP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of NAAP expression. Normal or standard values for NAAP expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to NAAP under conditions suitable for complex formation. The amount of standard complex formation may be

quantitated by various methods, such as photometric means. Quantities of NAAP expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, polynucleotides encoding NAAP may be used for  
5 diagnostic purposes. The polynucleotides which may be used include oligonucleotides, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of NAAP may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of NAAP, and to monitor regulation of NAAP levels during therapeutic intervention.

10 In one aspect, hybridization with PCR probes which are capable of detecting polynucleotides, including genomic sequences, encoding NAAP or closely related molecules may be used to identify nucleic acid sequences which encode NAAP. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe  
15 identifies only naturally occurring sequences encoding NAAP, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the NAAP encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:58-114 or from genomic sequences including promoters, enhancers, and introns of the NAAP gene.

20 Means for producing specific hybridization probes for polynucleotides encoding NAAP include the cloning of polynucleotides encoding NAAP or NAAP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter  
25 groups, for example, by radionuclides such as <sup>32</sup>P or <sup>35</sup>S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotides encoding NAAP may be used for the diagnosis of disorders associated with expression of NAAP. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed  
30 connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, a cancer of the adrenal

- gland, bladder, bone, bone marrow, brain, breast, cervix, colon, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease,
- 5 Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including
- 10 kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorder of the central nervous system, cerebral palsy, a neuroskeletal disorder, an autonomic nervous system disorder, a cranial nerve disorder, a spinal cord disease,
- 15 muscular dystrophy and other neuromuscular disorder, a peripheral nervous system disorder, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathy, myasthenia gravis, periodic paralysis, a mental disorder including mood, anxiety, and schizophrenic disorder, seasonal affective disorder (SAD), akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, and Tourette's disorder; a
- 20 developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism,
- 25 hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune
- 30 polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis,



glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic

5 lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; and an infection, such as those caused by a viral agent classified as adenovirus, arenavirus, bunyavirus, calicivirus, coronavirus, filovirus, hepadnavirus, herpesvirus, flavivirus, orthomyxovirus, parvovirus, papovavirus,

10 paramyxovirus, picornavirus, poxvirus, reovirus, retrovirus, rhabdovirus, or togavirus; an infection caused by a bacterial agent classified as pneumococcus, staphylococcus, streptococcus, bacillus, corynebacterium, clostridium, meningococcus, gonococcus, listeria, moraxella, kingella, haemophilus, legionella, bordetella, gram-negative enterobacterium including shigella, salmonella, or campylobacter, pseudomonas, vibrio, brucella, francisella, yersinia, bartonella, norcardium, actinomyces,

15 mycobacterium, spirochaetale, rickettsia, chlamydia, or mycoplasma; an infection caused by a fungal agent classified as aspergillus, blastomyces, dermatophytes, cryptococcus, coccidioides, malassezia, histoplasma, or other mycosis-causing fungal agent; and an infection caused by a parasite classified as plasmodium or malaria-causing, parasitic entamoeba, leishmania, trypanosoma, toxoplasma, pneumocystis carinii, intestinal protozoa such as giardia, trichomonas, tissue nematode such as

20 trichinella, intestinal nematode such as ascaris, lymphatic filarial nematode, trematode such as schistosoma, and cestode such as tapeworm. Polynucleotides encoding NAAP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered NAAP expression. Such qualitative or quantitative methods are well known

25 in the art.

In a particular embodiment, polynucleotides encoding NAAP may be used in assays that detect the presence of associated disorders, particularly those mentioned above. Polynucleotides complementary to sequences encoding NAAP may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization

30 complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of polynucleotides encoding NAAP

in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of NAAP, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding NAAP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease; or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier, thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding NAAP may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced *in vitro*. Oligomers will preferably contain a fragment of a polynucleotide encoding NAAP, or a fragment of a polynucleotide complementary to the polynucleotide encoding NAAP, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from polynucleotides encoding NAAP may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and

deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from polynucleotides encoding NAAP are used to amplify DNA using the polymerase chain reaction (PCR). The DNA  
5 may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSSCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence  
10 database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass  
15 spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

SNPs may be used to study the genetic basis of human disease. For example, at least 16 common SNPs have been associated with non-insulin-dependent diabetes mellitus. SNPs are also useful for examining differences in disease outcomes in monogenic disorders, such as cystic fibrosis,  
20 sickle cell anemia, or chronic granulomatous disease. For example, variants in the mannose-binding lectin, MBL2, have been shown to be correlated with deleterious pulmonary outcomes in cystic fibrosis. SNPs also have utility in pharmacogenomics, the identification of genetic variants that influence a patient's response to a drug, such as life-threatening toxicity. For example, a variation in N-acetyl transferase is associated with a high incidence of peripheral neuropathy in response to the  
25 anti-tuberculosis drug isoniazid, while a variation in the core promoter of the ALOX5 gene results in diminished clinical response to treatment with an anti-asthma drug that targets the 5-lipoxygenase pathway. Analysis of the distribution of SNPs in different populations is useful for investigating genetic drift, mutation, recombination, and selection, as well as for tracing the origins of populations and their migrations (Taylor, J.G. et al. (2001) Trends Mol. Med. 7:507-512; Kwok, P.-Y. and Z. Gu  
30 (1999) Mol. Med. Today 5:538-543; Nowotny, P. et al. (2001) Curr. Opin. Neurobiol. 11:637-641).

Methods which may also be used to quantify the expression of NAAP include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from

standard curves (Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236). The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid  
5 quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotides described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic  
10 variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment  
15 regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, NAAP, fragments of NAAP, or antibodies specific for NAAP may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

20 A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time (Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484;  
25 hereby expressly incorporated by reference herein). Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The  
30 resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression *in vivo*, as in the

case of a tissue or biopsy sample, or *in vitro*, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with *in vitro* model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) *Mol. Carcinog.* 24:153-159; Steiner, S. and N.L. Anderson (2000) *Toxicol. Lett.* 112-113:467-471). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity (see, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at <http://www.niehs.nih.gov/oc/news/toxchip.htm>). Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In an embodiment, the toxicity of a test compound can be assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another embodiment relates to the use of the polypeptides disclosed herein to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A

profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, *supra*). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of interest. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for NAAP to quantify the levels of NAAP expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) *Anal. Biochem.* 270:103-111; Mendoz, L.G. et al. (1999) *Biotechniques* 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) *Electrophoresis* 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated

biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art (Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662). Various types of microarrays are well known and thoroughly described in Schena, M., ed. (1999; DNA Microarrays: A Practical Approach, Oxford University Press, London).

In another embodiment of the invention, nucleic acid sequences encoding NAAP may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries (Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; Trask, B.J. (1991) Trends Genet. 7:149-154). Once mapped, the nucleic acid sequences may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP) (Lander, E.S. and D. Botstein (1986) Proc. Natl.

Acad. Sci. USA 83:7353-7357).

Fluorescent *in situ* hybridization (FISH) may be correlated with other physical and genetic map data (Heinz-Ulrich, et al. (1995) in Meyers, *supra*, pp. 965-968). Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM)

- 5 World Wide Web site. Correlation between the location of the gene encoding NAAP on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

*In situ* hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps.

- 10 Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any  
15 sequences mapping to that area may represent associated or regulatory genes for further investigation (Gatti, R.A. et al. (1988) Nature 336:577-580). The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

- In another embodiment of the invention, NAAP, its catalytic or immunogenic fragments, or  
20 oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between NAAP and the agent being tested may be measured.

- Another technique for drug screening provides for high throughput screening of compounds  
25 having suitable binding affinity to the protein of interest (Geysen, et al. (1984) PCT application WO84/03564). In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with NAAP, or fragments thereof, and washed. Bound NAAP is then detected by methods well known in the art. Purified NAAP can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively,  
30 non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding NAAP specifically compete with a test compound for binding NAAP.



In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with NAAP.

In additional embodiments, the nucleotide sequences which encode NAAP may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, including U.S. Ser. No. 60/343,004, U.S. Ser. No. 60/347,633, U.S. Ser. No. 60/359,498, and U.S. Ser. No. 60/351,749, are hereby expressly incorporated by reference.

## EXAMPLES

### I. Construction of cDNA Libraries

Incye cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incye Genomics, Palo Alto CA). Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Invitrogen), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERScript plasmid system (Invitrogen), using the

recommended procedures or similar methods known in the art (Ausubel et al., *supra*, ch. 5). Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Biosciences) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Invitrogen, Carlsbad CA), PCDNA2.1 plasmid (Invitrogen), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA plasmid (Invitrogen), PCMV-ICIS plasmid (Stratagene), p1GEN (Incyte Genomics, Palo Alto CA), pRARE (Incyte Genomics), or pINCY (Incyte Genomics), or derivatives thereof. Recombinant plasmids were transformed into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 $\alpha$ , DH10B, or ElectroMAX DH10B from Invitrogen.

## II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by *in vivo* excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

## III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the

MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Biosciences or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Amersham Biosciences); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (Ausubel et al., *supra*, ch. 7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM; PROTEOME databases with sequences from *Homo sapiens*, *Rattus norvegicus*, *Mus musculus*, *Caenorhabditis elegans*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, and *Candida albicans* (Incyte Genomics, Palo Alto CA); hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM (Haft, D.H. et al. (2001) Nucleic Acids Res. 29:41-43); and HMM-based protein domain databases such as SMART (Schultz, J. et al. (1998) Proc. Natl. Acad. Sci. USA 95:5857-5864; Letunic, I. et al. (2002) Nucleic Acids Res. 30:242-244). (HMM is a probabilistic approach which analyzes consensus primary structures of gene families; see, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide

sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, the PROTEOME databases, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM; and HMM-based protein domain databases such as SMART. Full length  
5 polynucleotide sequences are also analyzed using MACDNASIS PRO software (MiraiBio, Alameda CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

10 Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where  
15 applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID  
20 NO:58-114. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 2.

#### IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative nucleic acid-associated proteins were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is  
25 a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94; Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for  
30 Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode nucleic acid-associated proteins, the encoded polypeptides were analyzed by querying against PFAM models for nucleic acid-associated proteins. Potential nucleic acid-associated

proteins were also identified by homology to Incyte cDNA sequences that had been annotated as nucleic acid-associated proteins. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpi public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

#### **V. Assembly of Genomic Sequence Data with cDNA Sequence Data**

##### **"Stitched" Sequences**

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpi public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended

with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

#### **“Stretched” Sequences**

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public  
5 databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions  
10 may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore “stretched” or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

#### **15 VI. Chromosomal Mapping of NAAP Encoding Polynucleotides**

The sequences which were used to assemble SEQ ID NO:58-114 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched  
20 SEQ ID NO:58-114 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

25 Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome’s p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances  
30 are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI “GeneMap’99” World Wide Web site

(<http://www.ncbi.nlm.nih.gov/genemap/>), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

## VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound (Sambrook and Russell, *supra*, ch. 7; Ausubel et al., *supra*, ch. 4).

Analogous computer techniques applying BLAST were used to search for identical or related molecules in databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum} \{ \text{length}(\text{Seq. 1}), \text{length}(\text{Seq. 2}) \}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotides encoding NAAP are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived

from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding NAAP. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

#### VIII. Extension of NAAP Encoding Polynucleotides

Full length polynucleotides are produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing  $Mg^{2+}$ ,  $(NH_4)_2SO_4$ , and 2-mercaptoethanol, Taq DNA polymerase (Amersham Biosciences), ELONGASE enzyme (Invitrogen), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times;



Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 µl to 10 µl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Biosciences). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Biosciences), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Biosciences) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethylsulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Biosciences) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotides are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

#### **IX. Identification of Single Nucleotide Polymorphisms in NAAP Encoding**

### Polynucleotides

Common DNA sequence variants known as single nucleotide polymorphisms (SNPs) were identified in SEQ ID NO:58-114 using the LIFESEQ database (Incyte Genomics). Sequences from the same gene were clustered together and assembled as described in Example III, allowing the  
5 identification of all sequence variants in the gene. An algorithm consisting of a series of filters was used to distinguish SNPs from other sequence variants. Preliminary filters removed the majority of basecall errors by requiring a minimum Phred quality score of 15, and removed sequence alignment errors and errors resulting from improper trimming of vector sequences, chimeras, and splice variants. An automated procedure of advanced chromosome analysis analysed the original chromatogram files  
10 in the vicinity of the putative SNP. Clone error filters used statistically generated algorithms to identify errors introduced during laboratory processing, such as those caused by reverse transcriptase, polymerase, or somatic mutation. Clustering error filters used statistically generated algorithms to identify errors resulting from clustering of close homologs or pseudogenes, or due to contamination by non-human sequences. A final set of filters removed duplicates and SNPs found in immunoglobulins  
15 or T-cell receptors.

Certain SNPs were selected for further characterization by mass spectrometry using the high throughput MASSARRAY system (Sequenom, Inc.) to analyze allele frequencies at the SNP sites in four different human populations. The Caucasian population comprised 92 individuals (46 male, 46 female), including 83 from Utah, four French, three Venezuelan, and two Amish individuals. The  
20 African population comprised 194 individuals (97 male, 97 female), all African Americans. The Hispanic population comprised 324 individuals (162 male, 162 female), all Mexican Hispanic. The Asian population comprised 126 individuals (64 male, 62 female) with a reported parental breakdown of 43% Chinese, 31% Japanese, 13% Korean, 5% Vietnamese, and 8% other Asian. Allele frequencies were first analyzed in the Caucasian population; in some cases those SNPs which showed  
25 no allelic variance in this population were not further tested in the other three populations.

### X. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:58-114 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide  
30 fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250  $\mu$ Ci of [ $\gamma$ - $^{32}$ P] adenosine triphosphate (Amersham Biosciences), and T4 polynucleotide kinase (DuPont NEN,

Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Biosciences). An aliquot containing 10<sup>7</sup> counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

#### **XI. Microarrays**

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing; see, e.g., Baldeschweiler et al., *supra*), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Skena, M., ed. (1999) DNA Microarrays: A Practical Approach, Oxford University Press, London). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements (Skena, M. et al. (1995) *Science* 270:467-470; Shalon, D. et al. (1996) *Genome Res.* 6:639-645; Marshall, A. and J. Hodgson (1998) *Nat. Biotechnol.* 16:27-31).

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorption and mass spectrometry may be used for detection of hybridization. The degree of

complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

#### Tissue or Cell Sample Preparation

5           Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)<sup>+</sup> RNA is purified using the oligo-(dT) cellulose method. Each poly(A)<sup>+</sup> RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/ $\mu$ l oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/ $\mu$ l RNase inhibitor, 500  $\mu$ M dATP, 500  $\mu$ M dGTP, 500  $\mu$ M dTTP, 40  $\mu$ M dCTP, 40  $\mu$ M dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Biosciences). The reverse transcription  
10 reaction is performed in a 25 ml volume containing 200 ng poly(A)<sup>+</sup> RNA with GEMBRIGHT kits (Incyte Genomics). Specific control poly(A)<sup>+</sup> RNAs are synthesized by *in vitro* transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to stop the reaction and degrade the RNA. Samples are purified using two  
15 successive CHROMA SPIN 30 gel filtration spin columns (Clontech, Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14  $\mu$ l 5X SSC/0.2% SDS.

#### Microarray Preparation

20           Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5  $\mu$ g. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Biosciences).

25           Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C  
30 oven.

          Array elements are applied to the coated glass substrate using a procedure described in U.S. Patent No. 5,807,522, incorporated herein by reference. 1  $\mu$ l of the array element DNA, at an average

concentration of 100 ng/ $\mu$ l, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene).

Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water.

- 5 Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

### Hybridization

- Hybridization reactions contain 9  $\mu$ l of sample mixture consisting of 0.2  $\mu$ g each of Cy3 and  
10 Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm<sup>2</sup> coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140  $\mu$ l of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about  
15 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

### Detection

- Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines  
20 at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

- 25 In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is  
30 typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a

cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte Genomics). Array elements that exhibit at least about a two-fold change in expression, a signal-to-background ratio of at least about 2.5, and an element spot size of at least about 40%, are considered to be differentially expressed.

#### Expression

For example, both SEQ ID NO:62 and SEQ ID NO:71 showed differential expression in cancer cell lines or tumorous tissue versus non-cancerous cell lines or tissues, as determined by microarray analysis. The expression of SEQ ID NO:62 was decreased by at least two-fold in ovarian tumor tissue from a 79 year-old female donor as compared to matched normal ovarian tissue from the same donor. Matched normal and tumorigenic ovarian tissue samples were provided by the Huntsman Cancer Institute, (Salt Lake City, UT). Therefore, SEQ ID NO:62 is useful in diagnostic assays for ovarian cancer. The expression of SEQ ID NO:71 was increased by at least two-fold in lung squamous cell carcinoma tissue as compared to matched normal lung tissue from the same donor. Matched normal and tumorigenic lung tissue samples were provided by the Roy Castle International Centre for Lung Cancer Research (Liverpool, UK). Therefore, SEQ ID NO:71 is useful in diagnostic

assays for lung cancer.

In addition, the expression of SEQ ID NO:71 was increased at least two-fold in Tangier disease-derived fibroblasts compared to normal fibroblasts. Both types of cells were also cultured in the presence of cholesterol and compared with the same cell type cultured in the absence of cholesterol. Human fibroblasts were obtained from skin explants from both normal subjects and two patients with homozygous Tangier disease. Cell lines were immortalized by transfection with human papillomavirus 16 genes E6 and E7 and a neomycin resistance selectable marker. TD derived cells are deficient in an assay of apoA-I mediated tritiated cholesterol efflux. Therefore, SEQ ID NO:71 is also useful in diagnostic assays for Tangier disease.

In an alternative example, SEQ ID NO:79 and SEQ ID NO:81 were differentially expressed in human breast tumor cells. Histological and molecular evaluation of breast tumors reveals that the development of breast cancer evolves through a multi-step process whereby pre-malignant mammary epithelial cells undergo a relatively defined sequence of events leading to tumor formation. An early event in tumor development is ductal hyperplasia. Cells undergoing rapid neoplastic growth gradually progress to invasive carcinoma and become metastatic to the lung, bone, and potentially other organs. Several variables that may influence the process of tumor progression and malignant transformation include genetic factors, environmental factors, growth factors, and hormones. Based on the complexity of this process, it is critical to study a population of human mammary epithelial cells undergoing the process of malignant transformation, and to associate specific stages of progression with phenotypic and molecular characteristics. Primary breast epithelial cells (HMECs) were compared to various breast carcinoma lines at various stages of tumor progression.

Gene expression profiles of nonmalignant mammary epithelial cells were compared to gene expression profiles of various breast carcinoma lines at different stages of tumor progression. The cells were grown in defined serum-free H14 medium to 70-80% confluence prior to RNA harvest. Cell lines compared included: a) HMEC, a primary breast epithelial cell line isolated from a normal donor, b)MCF-10A, a breast mammary gland cell line isolated from a 36-year-old woman with fibrocystic breast disease, c)MCF7, a nonmalignant breast adenocarcinoma cell line isolated from the pleural effusion of a 69- year-old female, d)T-47D, a breast carcinoma cell line isolated from a pleural effusion obtained from a 54-year-old female with an infiltrating ductal carcinoma of the breast, e)Sk-BR-3, a breast adenocarcinoma cell line isolated from a malignant pleural effusion of a 43-year-old female, f)BT-20, a breast carcinoma cell line derived *in vitro* from cells emigrating out of thin slices of the tumor mass isolated from a 74-year-old female, g)MDA-mb-231, a breast tumor cell line isolated

from the pleural effusion of a 51-year-old female, and h)MDA-mb-435S, a spindle-shaped strain that evolved from the parent line (435) isolated by R. Cailleau from pleural effusion of a 31-year-old female with metastatic, ductal adenocarcinoma of the breast.

In various microarray experiments, both SEQ ID NO:79 and SEQ ID NO:81 were  
5 underexpressed by at least two-fold in the following tumor cell lines: MCF7 (nonmalignant breast adenocarcinoma), T47D (breast carcinoma), BT-20 (breast carcinoma), MDA-mb-231 (breast tumor cell line which also expresses the Wnt3 oncogene, EGF, and TGF-alpha), and MDA-mb-435S (spindle-shaped strain of breast tumor cell evolved from a metastatic, ductal adenocarcinoma cell of the breast).

10 These experiments indicate that SEQ ID NO:79 and SEQ ID NO:81 exhibited significant differential expression patterns using microarray techniques, and further establish their utility as diagnostic markers, disease staging or therapeutic agents which may be useful in a variety of conditions and diseases involving nucleic acid-associated proteins, including breast cancer.

In an alternative example, SEQ ID NO:90, SEQ ID NO:91, and SEQ ID NO:93 showed  
15 differential expression in tumorous or diseased tissue versus non-tumorous or healthy tissues, as determined by microarray analysis. Array elements that exhibited about at least a two-fold change in expression and a signal intensity over 250 units, a signal-to-background ratio of at least 2.5, and an element spot size of at least 40% were identified as differentially expressed using the GEMTOOLS program (Incyte Genomics).

20 For example, SEQ ID NO:90 showed increased expression in lung carcinoma tissue versus normal lung tissue as determined by microarray analysis. In one experiment, grossly uninvolved lung tissue from a 66 year-old male was compared to lung squamous cell carcinoma tissue from the same donor (Roy Castle International Centre for Lung Cancer Research, Liverpool, UK). In a second experiment, grossly uninvolved lung tissue from a 66 year-old female was compared to lung  
25 adenocarcinoma tissue from the same donor (Roy Castle International Centre for Lung Cancer Research, Liverpool, UK). Therefore, SEQ ID NO:90 is useful in monitoring treatment of, and diagnostic assays for, lung cancer and other cell proliferative disorders.

In another example, SEQ ID NO:90 showed differential expression in amygdala enterorhinal cortex tissue versus pooled brain tissue as determined by microarray analysis. Specific brain regions  
30 from 4 individual male donors (47, 48, 59, and 60 years old) were compared to a pooled brain control. The pooled brain control was reconstituted from the purified mRNA isolated from the major regions of the brain from two male brains (the 47-year-old male and the 48-year-old male). Therefore, SEQ ID



NO:90 is useful in monitoring treatment of, and diagnostic assays for, Alzheimer's disease and other neurological disorders.

In another example, SEQ ID NO:91 showed decreased expression in breast carcinoma cell lines versus a primary breast epithelial cell line isolated from a normal donor, as determined by  
5 microarray analysis. The gene expression profile of a nonmalignant mammary epithelial cell line was compared to the gene expression profiles of breast carcinoma lines at different stages of tumor progression. Cell lines compared included: a) BT-20, a breast carcinoma cell line derived *in vitro* from the cells emigrating out of thin slices of tumor mass isolated from a 74-year-old female, b) BT-474, a breast ductal carcinoma cell line that was isolated from a solid, invasive ductal carcinoma of the  
10 breast obtained from a 60-year-old woman, c) BT-483, a breast ductal carcinoma cell line that was isolated from a papillary invasive ductal tumor obtained from a 23-year-old normal, menstruating, parous female with a family history of breast cancer, d) Hs 578T, a breast ductal carcinoma cell line isolated from a 74-year-old female with breast carcinoma, e) MCF7, a nonmalignant breast adenocarcinoma cell line isolated from the pleural effusion of a 69-year-old female, f) MCF-10A, a  
15 breast mammary gland (luminal ductal characteristics) cell line isolated from a 36-year-old woman with fibrocystic breast disease, g) MDA-MB-468, a breast adenocarcinoma cell line isolated from the pleural effusion of a 51-year-old female with metastatic adenocarcinoma of the breast, and h) HMEC, a primary breast epithelial cell line isolated from a normal donor. SEQ ID NO:91 showed decreased expression in lines a) through e) above. Therefore, SEQ ID NO:91 is useful in monitoring treatment  
20 of, and diagnostic assays for, breast cancer and other cell proliferative disorders.

In another example, SEQ ID NO:93 showed decreased expression in breast carcinoma cells treated with estradiol versus untreated breast carcinoma cells as determined by microarray analysis. BT-20 is a breast carcinoma cell line derived *in vitro* from the cells emigrating out thin slices of the tumor mass isolated from a 74-year-old female. BT-20 cells were treated with  $\alpha$ -estradiol for 4, 8, 14,  
25 24, 36, and 48 hours. These treated cells were compared to untreated BT-20 cells kept in culture for the same amount of time. Therefore, SEQ ID NO:93 is useful in monitoring treatment of, and diagnostic assays for, breast cancer and other cell proliferative disorders.

In another example, SEQ ID NO:93 showed increased expression in coronary artery epithelial cells versus pooled vascular endothelial cells as determined by microarray analysis. Human vascular  
30 and microvascular endothelial cells were isolated from coronary artery, umbilical artery and vein, uterine microvasculature, iliac artery, dermal microvasculature, pulmonary artery, aorta. All cell types in this experiment are used as experimental models for investigating the role of endothelium in human

vascular biology. Therefore, SEQ ID NO:93 is useful in monitoring treatment of, and diagnostic assays for, cardiovascular and other autoimmune/inflammation disorders.

In an alternative example, SEQ ID NO:101, SEQ ID NO:103, SEQ ID NO:106, SEQ ID NO:112, and SEQ ID NO:113 showed differential expression in tumorous tissues versus non-tumorous tissues or in treated versus untreated cell lines, as determined by microarray analysis. Array elements that exhibited about at least a two-fold change in expression and a signal intensity over 250 units, a signal-to-background ratio of at least 2.0, and an element spot size of at least 40% were identified as differentially expressed using the GEMTOOLS program (Incyte Genomics).

For example, the expression of SEQ ID NO:101 was decreased at least two-fold in an ovarian adenocarcinoma when matched with normal tissue from the same donor. The tumorous ovary tissue was obtained from ovarian adenocarcinoma from a 79-year-old female. Normal ovary tissue was obtained from ovary from the same donor. Therefore, SEQ ID NO:101 is useful in diagnostic assays for ovarian adenocarcinoma. Matched normal and tumorigenic ovary tissue samples were provided by the Huntsman Cancer Institute, (Salt Lake City, UT).

The gene expression profile of a nonmalignant mammary epithelial cell line (MCF-10A) or a nonmalignant mammary epithelial cell line (HMEC) was compared to the gene expression profiles of breast carcinoma lines at different stages of tumor progression. Cell lines compared included: a) HMEC, a primary breast epithelial cell line isolated from a normal donor, b) MCF-10A, a breast mammary gland cell line isolated from a 36-year-old woman with fibrocystic breast disease; c) MCF7, a nonmalignant breast adenocarcinoma cell line isolated from the pleural effusion of a 69-year-old female; d) T-47D, a breast carcinoma cell line isolated from a pleural effusion obtained from a 54-year-old female with an infiltrating ductal carcinoma of the breast; e) Sk-BR-3, a breast adenocarcinoma cell line isolated from a malignant pleural effusion of a 43-year-old female; f) BT-20, a breast carcinoma cell line derived *in vitro* from tumor mass isolated from a 74-year-old female; g) MDA-mb-231, a breast tumor cell line isolated from the pleural effusion of a 51-year old female; and h) MDA-mb-435S, a spindle shaped strain that evolved from the parent line (435) isolated from the pleural effusion of a 31-year-old female with metastatic, ductal adenocarcinoma of the breast. All cell lines were grown in the supplier's recommended medium to 70-80% confluence prior to comparison. The expression of SEQ ID NO:101 was decreased at least two-fold in three (MCF7, T-47D, and MDA-mb-231) of the seven cell lines when compared with HMEC or MCF-10A cell lines. These experiments indicate that SEQ ID NO:101 was significantly under-expressed in the breast tumor cell lines tested, establishing the utility of SEQ ID NO:101 as a diagnostic marker for disease

staging or as a potential therapeutic target for breast cancer.

In another example, SEQ ID NO:103, exhibited differential expression by microarray analysis. Early confluent C3A cells were treated with mifepristone, progesterone, beclomethasone, medroxyprogesterone, budesonide, prednisone, dexamethasone, betamethasone, or danazol at concentrations of 1  $\mu$ M, 10  $\mu$ M, and 100  $\mu$ M for 1, 3, and 6 hours. In all cases mRNA from untreated early confluent C3A cells were prepared in parallel as described below. The expression of SEQ ID NO:103 was decreased at least two-fold in HCA3 cells when treated with Progesterone, Beclomethasone, Medroxyprogesterone, Budesonide, Prednisone, Dexamethasone, and Betamethasone at doses between 1  $\mu$ M and 100  $\mu$ M for between 1 and 6 hours, and with Catechol plus water at 100  $\mu$ M for 1, 3, and 6 hours. Therefore, SEQ ID NO:103 is useful as a diagnostic marker for disease staging or as a potential therapeutic target for liver disorders such as adenomatosis, cholestasis, cirrhosis, hemangioma, Henoch-Schonlein purpura, hepatitis, hepatocellular and metastatic carcinomas, idiopathic thrombocytopenic purpura, porphyria, sarcoidosis, and Wilson disease.

In another example, the expression of SEQ ID NO:106 was increased at least four-fold in lung squamous cell carcinoma when matched with normal tissue from the same donor. The tumorous tissue was obtained from lung squamous cell carcinoma from a 68-year-old female. Normal lung tissue was obtained from the same donor. Therefore, SEQ ID NO:106 is useful in diagnostic assays for lung squamous cell carcinoma, in disease staging and as a potential therapeutic target. Matched normal and tumorigenic lung tissue samples are provided by the Roy Castle International Centre for Lung Cancer Research (Liverpool UK). Further, the expression of SEQ ID NO:106 was decreased at least three-fold in colon tumor when matched with normal tissue from the same donor. Tumorous colon tissue was obtained from a 67-year-old donor with moderately differentiated colon adenocarcinoma. Normal tissue was obtained from the colon of the same donor. Therefore, SEQ ID NO:106 is useful as a diagnostic marker, for disease staging, or as a potential therapeutic target for colon adenocarcinoma. Matched normal and tumorigenic colon tissue samples were provided by the Huntsman Cancer Institute (Salt Lake City, UT).

In another example, the expression of SEQ ID NO:112 was decreased at least two-fold in two colon adenocarcinomas when matched with normal tissue from the same donor. The tumorous tissue was obtained from a 64-year-old female with moderately differentiated colon adenocarcinoma and from an 83-year old female with colon adenocarcinoma. Normal tissues were obtained from the colons of the same donors. Therefore, SEQ ID NO:112 is useful as a diagnostic marker, for disease staging, or as a potential therapeutic target for colon adenocarcinoma. Matched normal and

tumorigenic colon tissue samples are provided by the Huntsman Cancer Institute (Salt Lake City, UT).

In yet another example, SEQ ID NO:113 showed differential expression in inflammatory responses as determined by microarray analysis. The expression of SEQ ID NO:113 was increased by at least two-fold in three cell lines, Daudi (a B lymphoblast cell line [Burkitt's lymphoma]), Jurkat (an acute T cell leukemia cell line that grows in the absence of external stimuli), and THP-1 (a promonocyte cell line isolated from the peripheral blood of a 1-year-old male with acute monocytic leukemia). These three cell lines are used extensively for the study of signaling in human T cells and B cells. These cell lines were treated with various concentrations of PMA (a broad activator of protein kinase C-dependent (PKC) pathways), with or without Ionomycin (a calcium ionophore that permits the entry of calcium in the cell), with or without Brefeldin A (an antiviral antibiotic factor produced by Penicillium brefeldianum), with or without antibodies to CD3 or CD8, or with or without LPS (lipopolysaccharide). The expression of SEQ ID NO:113 was increased by at least two-fold in three tumorous cell lines from breast (MDA-mb-231, T-47D; and MCF7) when treated with 1 or 10 ng/ml of TNF $\alpha$ . See descriptions supra. The expression of SEQ ID NO:113 was also increased by at least 2.9-fold in DU 145, a prostate carcinoma cell line isolated from a metastatic site in the brain of a 69-year-old male with widespread metastatic prostate carcinoma, when compared to Prec, a primary prostate epithelial cell line isolated from a normal donor; and increased by at least two-fold in the following endothelial cell lines: ECV304 (a cell line derived from the endothelium of the human umbilical vein) treated with PMA and Ionomycin or TNF $\alpha$ ; HIAEC (primary cells derived from the endothelium of an iliac artery); HMVEC (primary cells derived from microvascular of human skin endothelial cells); HUAEC (primary cells derived from the endothelium of an umbilical artery); and UtMVECMyo (primary cells derived from the uterine myometrium microvasculature) all treated with 10 ng/ml TNF- $\alpha$  when compared to untreated cells of the same cell line; and HUVEC (primary cells derived from the endothelium of the human umbilical vein) treated with TNF $\alpha$  and cycloheximide when compared to untreated HUVEC cells treated with cycloheximide.

Further, the expression of SEQ ID NO:113 was decreased by at least two-fold in HepG2 (a human hepatoma cell line isolated from a 15-year-old male with a liver tumor) when treated with TNF $\alpha$ . SEQ ID NO:113 was decreased at least two-fold in five out of six breast tumor cell lines (BT20, HMEC, MDA-mb-231, MDA-mb-435S, T-47D; see descriptions supra) when compared with MCF10A cell line. The expression of SEQ ID NO:113 was decreased at least 2.5-fold in a colon cancer when matched with normal tissue from the same donor. The tumorous colon tissue was obtained from a 73-year-old female with colon cancer. Normal colon tissue was obtained from the

colon of the same donor. Also, the expression of SEQ ID NO:113 was decreased at least two-fold in a donor with lung adenocarcinoma and decreased at least 2.5-fold in four donors with lung squamous cell carcinoma. The adenocarcinoma tissue was obtained from a 66-year-old female with lung adenocarcinoma. The squamous cell carcinoma tissue was obtained from a 73-year-old male, a 68-year-old female, a 66-year-old male, and a 73-year-old male. Normal lung tissues were obtained from the lungs of the same donors. All matched normal and tumorous lung tissue samples were provided by the Roy Castle International Centre for Lung Cancer Research (Liverpool UK). Therefore, SEQ ID NO:113 is useful in monitoring treatment of, and diagnostic assays for, autoimmune/inflammation disorders, breast cancer, colon cancer, and lung cancer.

## 10 XII. Complementary Polynucleotides

Sequences complementary to the NAAP-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring NAAP. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of NAAP. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the NAAP-encoding transcript.

## XIII. Expression of NAAP

20 Expression and purification of NAAP is achieved using bacterial or virus-based expression systems. For expression of NAAP in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express NAAP upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of NAAP in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant *Autographica californica* nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding NAAP by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to

infect *Spodoptera frugiperda* (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus (Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945).

- 5           In most expression systems, NAAP is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from *Schistosoma japonicum*, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Biosciences).
- 10       Following purification, the GST moiety can be proteolytically cleaved from NAAP at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel et al. (*supra*, ch. 10 and 16).
- 15       Purified NAAP obtained by these methods can be used directly in the assays shown in Examples XVII, XVIII, and XIX, where applicable.

#### XIV. Functional Assays

- NAAP function is assessed by expressing the sequences encoding NAAP at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression
- 20       vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT plasmid (Invitrogen, Carlsbad CA) and PCR3.1 plasmid (Invitrogen), both of which contain the cytomegalovirus promoter. 5-10  $\mu$ g of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2  $\mu$ g of an additional plasmid containing sequences
- 25       encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to
- 30       evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium

iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994; Flow Cytometry, Oxford, New York NY).

The influence of NAAP on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding NAAP and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding NAAP and other genes of interest can be analyzed by northern analysis or microarray techniques.

#### 15 XV. Production of NAAP Specific Antibodies

NAAP substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to immunize animals (e.g., rabbits, mice, etc.) and to produce antibodies using standard protocols.

Alternatively, the NAAP amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art (Ausubel et al., *supra*, ch. 11).

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using Fmoc chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity (Ausubel et al., *supra*). Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-NAAP activity by, for example, binding the peptide or NAAP to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

#### XVI. Purification of Naturally Occurring NAAP Using Specific Antibodies

Naturally occurring or recombinant NAAP is substantially purified by immunoaffinity

chromatography using antibodies specific for NAAP. An immunoaffinity column is constructed by covalently coupling anti-NAAP antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Biosciences). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

- 5 Media containing NAAP are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of NAAP (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/NAAP binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and NAAP is collected.

#### 10 XVII. Identification of Molecules Which Interact with NAAP

- NAAP, or biologically active fragments thereof, are labeled with  $^{125}\text{I}$  Bolton-Hunter reagent (Bolton, A.E. and W.M. Hunter (1973) *Biochem. J.* 133:529-539). Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled NAAP, washed, and any wells with labeled NAAP complex are assayed. Data obtained using different concentrations of NAAP are  
15 used to calculate values for the number, affinity, and association of NAAP with the candidate molecules.

Alternatively, molecules interacting with NAAP are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989; *Nature* 340:245-246), or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

- 20 NAAP may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

#### XVIII. Demonstration of NAAP Activity

- 25 NAAP activity is measured by its ability to stimulate transcription of a reporter gene (Liu, H.Y. et al. (1997) *EMBO J.* 16:5289-5298). The assay entails the use of a well characterized reporter gene construct, LexA<sub>op</sub>-LacZ, that consists of LexA DNA transcriptional control elements (LexA<sub>op</sub>) fused to sequences encoding the E. coli LacZ enzyme. The methods for constructing and expressing fusion genes, introducing them into cells, and measuring LacZ enzyme activity, are well  
30 known to those skilled in the art. Sequences encoding NAAP are cloned into a plasmid that directs the synthesis of a fusion protein, LexA-NAAP, consisting of NAAP and a DNA binding domain derived from the LexA transcription factor. The resulting plasmid, encoding a LexA-NAAP fusion



protein, is introduced into yeast cells along with a plasmid containing the LexA<sub>op</sub>-LacZ reporter gene. The amount of LacZ enzyme activity associated with LexA-NAAP transfected cells, relative to control cells, is proportional to the amount of transcription stimulated by the NAAP.

Alternatively, NAAP activity is measured by its ability to bind zinc. A 5-10  $\mu$ M sample  
5 solution in 2.5 mM ammonium acetate solution at pH 7.4 is combined with 0.05 M zinc sulfate solution (Aldrich, Milwaukee WI) in the presence of 100  $\mu$ M dithiothreitol with 10% methanol added. The sample and zinc sulfate solutions are allowed to incubate for 20 minutes. The reaction solution is passed through a VYDAC column (Grace Vydac, Hesperia, CA) with approximately 300 Angstrom bore size and 5  $\mu$ M particle size to isolate zinc-sample complex from the solution, and into a mass  
10 spectrometer (PE Sciex, Ontario, Canada). Zinc bound to sample is quantified using the functional atomic mass of 63.5 Da observed by Whittall, R. M. et al. ((2000) Biochemistry 39:8406-8417).

In the alternative, a method to determine nucleic acid binding activity of NAAP involves a polyacrylamide gel mobility-shift assay. In preparation for this assay, NAAP is expressed by transforming a mammalian cell line such as COS7, HeLa or CHO with a eukaryotic expression vector  
15 containing NAAP cDNA. The cells are incubated for 48-72 hours after transformation under conditions appropriate for the cell line to allow expression and accumulation of NAAP. Extracts containing solubilized proteins can be prepared from cells expressing NAAP by methods well known in the art. Portions of the extract containing NAAP are added to [<sup>32</sup>P]-labeled RNA or DNA. Radioactive nucleic acid can be synthesized *in vitro* by techniques well known in the art. The  
20 mixtures are incubated at 25°C in the presence of RNase- and DNase-inhibitors under buffered conditions for 5-10 minutes. After incubation, the samples are analyzed by polyacrylamide gel electrophoresis followed by autoradiography. The presence of a band on the autoradiogram indicates the formation of a complex between NAAP and the radioactive transcript. A band of similar mobility will not be present in samples prepared using control extracts prepared from untransformed cells.

25 In the alternative, a method to determine methylase activity of NAAP measures transfer of radiolabeled methyl groups between a donor substrate and an acceptor substrate. Reaction mixtures (50  $\mu$ l final volume) contain 15 mM HEPES, pH 7.9, 1.5 mM MgCl<sub>2</sub>, 10 mM dithiothreitol, 3% polyvinylalcohol, 1.5  $\mu$ Ci [*methyl*-<sup>3</sup>H]AdoMet (0.375  $\mu$ M AdoMet) (DuPont-NEN), 0.6  $\mu$ g NAAP, and acceptor substrate (e.g., 0.4  $\mu$ g [<sup>35</sup>S]RNA, or 6-mercaptopurine (6-MP) to 1 mM final concentration).  
30 Reaction mixtures are incubated at 30°C for 30 minutes, then 65°C for 5 minutes.

Analysis of [*methyl*-<sup>3</sup>H]RNA is as follows: (1) 50  $\mu$ l of 2 x loading buffer (20 mM Tris-HCl, pH 7.6, 1 M LiCl, 1 mM EDTA, 1% sodium dodecyl sulphate (SDS)) and 50  $\mu$ l oligo d(T)-cellulose

(10 mg/ml in 1 x loading buffer) are added to the reaction mixture, and incubated at ambient temperature with shaking for 30 minutes. (2) Reaction mixtures are transferred to a 96-well filtration plate attached to a vacuum apparatus. (3) Each sample is washed sequentially with three 2.4 ml aliquots of 1 x oligo d(T) loading buffer containing 0.5% SDS, 0.1% SDS, or no SDS. (4) RNA is  
5 eluted with 300  $\mu$ l of water into a 96-well collection plate, transferred to scintillation vials containing liquid scintillant, and radioactivity determined.

Analysis of [*methyl*- $^3$ H]6-MP is as follows: (1) 500  $\mu$ l 0.5 M borate buffer, pH 10.0, and then 2.5 ml of 20% (v/v) isoamyl alcohol in toluene are added to the reaction mixtures. (2) The samples are mixed by vigorous vortexing for ten seconds. (3) After centrifugation at 700g for 10 minutes, 1.5  
10 ml of the organic phase is transferred to scintillation vials containing 0.5 ml absolute ethanol and liquid scintillant, and radioactivity determined. (4) Results are corrected for the extraction of 6-MP into the organic phase (approximately 41%).

In the alternative, type I topoisomerase activity of NAAP can be assayed based on the relaxation of a supercoiled DNA substrate. NAAP is incubated with its substrate in a buffer lacking  
15  $Mg^{2+}$  and ATP, the reaction is terminated, and the products are loaded on an agarose gel. Altered topoisomers can be distinguished from supercoiled substrate electrophoretically. This assay is specific for type I topoisomerase activity because  $Mg^{2+}$  and ATP are necessary cofactors for type II topoisomerases.

Type II topoisomerase activity of NAAP can be assayed based on the decatenation of a  
20 kinetoplast DNA (KDNA) substrate. NAAP is incubated with KDNA, the reaction is terminated, and the products are loaded on an agarose gel. Monomeric circular KDNA can be distinguished from catenated KDNA electrophoretically. Kits for measuring type I and type II topoisomerase activities are available commercially from Topogen (Columbus OH).

ATP-dependent RNA helicase unwinding activity of NAAP can be measured by the method  
25 described by Zhang and Grosse (1994; Biochemistry 33:3906-3912). The substrate for RNA unwinding consists of  $^{32}$ P-labeled RNA composed of two RNA strands of 194 and 130 nucleotides in length containing a duplex region of 17 base-pairs. The RNA substrate is incubated together with ATP,  $Mg^{2+}$ , and varying amounts of NAAP in a Tris-HCl buffer, pH 7.5, at 37°C for 30 minutes. The single-stranded RNA product is then separated from the double-stranded RNA substrate by  
30 electrophoresis through a 10% SDS-polyacrylamide gel, and quantitated by autoradiography. The amount of single-stranded RNA recovered is proportional to the amount of NAAP in the preparation.

In the alternative, NAAP function is assessed by expressing the sequences encoding NAAP

at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include pCMV SPORT (Life Technologies) and pCR3.1 (Invitrogen Corporation, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10  $\mu$ g of recombinant vector are transiently transfected into a human cell line, preferably of endothelial or hematopoietic origin, using either liposome formulations or electroporation. 1-2  $\mu$ g of an additional plasmid containing sequences encoding a marker protein are co-transfected.

Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; CLONTECH), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties.

FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M. G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of NAAP on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding NAAP and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Inc., Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding NAAP and other genes of interest can be analyzed by northern analysis or microarray techniques.

Pseudouridine synthase activity of NAAP is assayed using a tritium ( $^3\text{H}$ ) release assay modified from Nurse et al. ((1995) RNA 1:102-112), which measures the release of  $^3\text{H}$  from the C<sub>5</sub>

position of the pyrimidine component of uridylylate (U) when  $^3\text{H}$ -radiolabeled U in RNA is isomerized to pseudouridine ( $\psi$ ). A typical 500  $\mu\text{l}$  assay mixture contains 50 mM HEPES buffer (pH 7.5), 100 mM ammonium acetate, 5 mM dithiothreitol, 1 mM EDTA, 30 units RNase inhibitor, and 0.1-4.2  $\mu\text{M}$   $[5\text{-}^3\text{H}]\text{tRNA}$  (approximately 1  $\mu\text{Ci/nmol}$  tRNA). The reaction is initiated by the addition of <5  $\mu\text{l}$  of a concentrated solution of NAAP (or sample containing NAAP) and incubated for 5 min at 37  $^{\circ}\text{C}$ . Portions of the reaction mixture are removed at various times (up to 30 min) following the addition of NAAP and quenched by dilution into 1 ml 0.1 M HCl containing Norit-SA3 (12% w/v). The quenched reaction mixtures are centrifuged for 5 min at maximum speed in a microcentrifuge, and the supernatants are filtered through a plug of glass wool. The pellet is washed twice by resuspension in 1 ml 0.1 M HCl, followed by centrifugation. The supernatants from the washes are separately passed through the glass wool plug and combined with the original filtrate. A portion of the combined filtrate is mixed with scintillation fluid (up to 10 ml) and counted using a scintillation counter. The amount of  $^3\text{H}$  released from the RNA and present in the soluble filtrate is proportional to the amount of pseudouridine synthase activity in the sample (Ramamurthy, V. (1999) J. Biol. Chem. 274:22225-22230).

In the alternative, pseudouridine synthase activity of NAAP is assayed at 30  $^{\circ}\text{C}$  to 37  $^{\circ}\text{C}$  in a mixture containing 100 mM Tris-HCl (pH 8.0), 100 mM ammonium acetate, 5 mM  $\text{MgCl}_2$ , 2 mM dithiothreitol, 0.1 mM EDTA, and 1-2 fmol of  $[^{32}\text{P}]$ -radiolabeled runoff transcripts (generated *in vitro* by an appropriate RNA polymerase, i.e., T7 or SP6) as substrates. NAAP is added to initiate the reaction or omitted from the reaction in control samples. Following incubation, the RNA is extracted with phenol-chloroform, precipitated in ethanol, and hydrolyzed completely to 3-nucleotide monophosphates using RNase  $\text{T}_2$ . The hydrolysates are analyzed by two-dimensional thin layer chromatography, and the amount of  $^{32}\text{P}$  radiolabel present in the  $\psi\text{MP}$  and UMP spots are evaluated after exposing the thin layer chromatography plates to film or a PhosphorImager screen. Taking into account the relative number of uridylylate residues in the substrate RNA, the relative amount  $\psi\text{MP}$  and UMP are determined and used to calculate the relative amount of  $\psi$  per tRNA molecule (expressed in mol  $\psi$  /mol of tRNA or mol  $\psi$  /mol of tRNA/minute), which corresponds to the amount of pseudouridine synthase activity in the NAAP sample (Lecoite, F. et al. (1998) J. Biol. Chem. 273:1316-1323).

$\text{N}^2,\text{N}^2$ -dimethylguanosine transferase ( $(\text{m}^2_2\text{G})$ methyltransferase) activity of NAAP is measured in a 160  $\mu\text{l}$  reaction mixture containing 100 mM Tris-HCl (pH 7.5), 0.1 mM EDTA, 10 mM  $\text{MgCl}_2$ , 20 mM  $\text{NH}_4\text{Cl}$ , 1mM dithiothreitol, 6.2  $\mu\text{M}$  *S*-adenosyl-L- $[methyl\text{-}^3\text{H}]\text{methionine}$  (30-70

Ci/mM), 8  $\mu\text{g m}^2\text{G}$ -deficient tRNA or wild type tRNA from yeast, and approximately 100  $\mu\text{g}$  of purified NAAP or a sample comprising NAAP. The reactions are incubated at 30 °C for 90 min and chilled on ice. A portion of each reaction is diluted to 1 ml in water containing 100  $\mu\text{g}$  BSA. 1 ml of 2 M HCl is added to each sample and the acid insoluble products are allowed to precipitate on ice for 20 min before being collected by filtration through glass fiber filters. The collected material is washed several times with HCl and quantitated using a liquid scintillation counter. The amount of  $^3\text{H}$  incorporated into the  $\text{m}^2\text{G}$ -deficient, acid-insoluble tRNAs is proportional to the amount of  $\text{N}^2\text{,N}^2$ -dimethylguanosine transferase activity in the NAAP sample. Reactions comprising no substrate tRNAs, or wild-type tRNAs that have already been modified, serve as control reactions which should not yield acid-insoluble  $^3\text{H}$ -labeled products.

Polyadenylation activity of NAAP is measured using an *in vitro* polyadenylation reaction. The reaction mixture is assembled on ice and comprises 10  $\mu\text{l}$  of 5 mM dithiothreitol, 0.025% (v/v) NONIDET P-40, 50 mM creatine phosphate, 6.5% (w/v) polyvinyl alcohol, 0.5 unit/ $\mu\text{l}$  RNAGUARD (Pharmacia), 0.025  $\mu\text{g}/\mu\text{l}$  creatine kinase, 1.25 mM cordycepin 5'-triphosphate, and 3.75 mM  $\text{MgCl}_2$ , in a total volume of 25  $\mu\text{l}$ . 60 fmol of CstF, 50 fmol of CPSF, 240 fmol of PAP, 4  $\mu\text{l}$  of crude or partially purified CF II and various amounts of amounts CF I are then added to the reaction mix. The volume is adjusted to 23.5  $\mu\text{l}$  with a buffer containing 50 mM TrisHCl, pH 7.9, 10% (v/v) glycerol, and 0.1 mM Na-EDTA. The final ammonium sulfate concentration should be below 20 mM. The reaction is initiated (on ice) by the addition of 15 fmol of  $^{32}\text{P}$ -labeled pre-mRNA template, along with 2.5  $\mu\text{g}$  of unlabeled tRNA, in 1.5  $\mu\text{l}$  of water. Reactions are then incubated at 30 °C for 75-90 min and stopped by the addition of 75  $\mu\text{l}$  (approximately two-volumes) of proteinase K mix (0.2 M Tris-HCl, pH 7.9, 300 mM NaCl, 25 mM Na-EDTA, 2% (w/v) SDS), 1  $\mu\text{l}$  of 10 mg/ml proteinase K, 0.25  $\mu\text{l}$  of 20 mg/ml glycogen, and 23.75  $\mu\text{l}$  of water). Following incubation, the RNA is precipitated with ethanol and analyzed on a 6% (w/v) polyacrylamide, 8.3 M urea sequencing gel. The dried gel is developed by autoradiography or using a phosphoimager. Cleavage activity is determined by comparing the amount of cleavage product to the amount of pre-mRNA template. The omission of any of the polypeptide components of the reaction and substitution of NAAP is useful for identifying the specific biological function of NAAP in pre-mRNA polyadenylation (Rüegsegger, U. et al. (1996) J. Biol. Chem. 271:6107-6113; and references within).

tRNA synthetase activity is measured as the aminoacylation of a substrate tRNA in the presence of [ $^{14}\text{C}$ ]-labeled amino acid. NAAP is incubated with [ $^{14}\text{C}$ ]-labeled amino acid and the appropriate cognate tRNA (for example, [ $^{14}\text{C}$ ]alanine and tRNA<sup>ala</sup>) in a buffered solution.  $^{14}\text{C}$ -

labeled product is separated from free [ $^{14}\text{C}$ ]amino acid by chromatography, and the incorporated  $^{14}\text{C}$  is quantified by scintillation counter. The amount of  $^{14}\text{C}$ -labeled product detected is proportional to the activity of NAAP in this assay.

In the alternative, NAAP activity is measured by incubating a sample containing NAAP in a solution containing 1 mM ATP, 5 mM Hepes-KOH (pH 7.0), 2.5 mM KCl, 1.5 mM magnesium chloride, and 0.5 mM DTT along with misacylated [ $^{14}\text{C}$ ]-Glu-tRNA<sup>Gln</sup> (e.g., 1  $\mu\text{M}$ ) and a similar concentration of unlabeled L-glutamine. Following the quenching of the reaction with 3 M sodium acetate (pH 5.0), the mixture is extracted with an equal volume of water-saturated phenol, and the aqueous and organic phases are separated by centrifugation at  $15,000 \times g$  at room temperature for 1 min. The aqueous phase is removed and precipitated with 3 volumes of ethanol at  $-70^\circ\text{C}$  for 15 min. The precipitated aminoacyl-tRNAs are recovered by centrifugation at  $15,000 \times g$  at  $4^\circ\text{C}$  for 15 min. The pellet is resuspended in of 25 mM KOH, deacylated at  $65^\circ\text{C}$  for 10 min., neutralized with 0.1 M HCl (to final pH 6-7), and dried under vacuum. The dried pellet is resuspended in water and spotted onto a cellulose TLC plate. The plate is developed in either isopropanol/formic acid/water or ammonia/water/chloroform/ methanol. The image is subjected to densitometric analysis and the relative amounts of Glu and Gln are calculated based on the  $R_f$  values and relative intensities of the spots. NAAP activity is calculated based on the amount of Gln resulting from the transformation of Glu while acylated as Glu-tRNA<sup>Gln</sup> (adapted from Curnow, A.W. et al. (1997) Proc. Natl. Acad. Sci. USA 94:11819-26).

#### 20 XIX. Identification of NAAP Agonists and Antagonists

Agonists or antagonists of NAAP activation or inhibition may be tested using the assays described in section XVIII. Agonists cause an increase in NAAP activity and antagonists cause a decrease in NAAP activity.

25 Various modifications and variations of the described compositions, methods, and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. It will be appreciated that the invention provides novel and useful proteins, and their encoding polynucleotides, which can be used in the drug discovery process, as well as methods for using these compositions for the detection, diagnosis, and treatment of diseases and conditions.

30 Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Nor should the description of such embodiments be considered exhaustive or limit the invention to the

precise forms disclosed. Furthermore, elements from one embodiment can be readily recombined with elements from one or more other embodiments. Such combinations can form a number of embodiments within the scope of the invention. It is intended that the scope of the invention be defined by the following claims and their equivalents.

Table 1

IncYTE Project ID	Polypeptide SEQ ID NO:	IncYTE Polypeptide ID	Polynucleotide SEQ ID NO:	IncYTE Polynucleotide ID	IncYTE Full Length Clones
7506140	1	7506140CD1	58	7506140CB1	
1889415	2	1889415CD1	59	1889415CB1	
7506047	3	7506047CD1	60	7506047CB1	
7505849	4	7505849CD1	61	7505849CB1	90179245CA2, 90179313CA2
7505972	5	7505972CD1	62	7505972CB1	6717651CA2
7505991	6	7505991CD1	63	7505991CB1	
7506003	7	7506003CD1	64	7506003CB1	
6483977	8	6483977CD1	65	6483977CB1	8692157CA2
6301777	9	6301777CD1	66	6301777CB1	6301777CA2
7505976	10	7505976CD1	67	7505976CB1	
7506016	11	7506016CD1	68	7506016CB1	
7506086	12	7506086CD1	69	7506086CB1	
4176657	13	4176657CD1	70	4176657CB1	
7506056	14	7506056CD1	71	7506056CB1	
7506185	15	7506185CD1	72	7506185CB1	
8096611	16	8096611CD1	73	8096611CB1	
8174603	17	8174603CD1	74	8174603CB1	
3101042	18	3101042CD1	75	3101042CB1	
4972035	19	4972035CD1	76	4972035CB1	
7506265	20	7506265CD1	77	7506265CB1	
7506304	21	7506304CD1	78	7506304CB1	
7506198	22	7506198CD1	79	7506198CB1	90097138CA2
1381261	23	1381261CD1	80	1381261CB1	
6803876	24	6803876CD1	81	6803876CB1	5386286CA2
7506281	25	7506281CD1	82	7506281CB1	90070544CA2
7506175	26	7506175CD1	83	7506175CB1	
7506303	27	7506303CD1	84	7506303CB1	4996649CA2
7353336	28	7353336CD1	85	7353336CB1	



Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Incyte Full Length Clones
3001652	29	3001652CD1	86	3001652CB1	
1689128	30	1689128CD1	87	1689128CB1	
2362969	31	2362969CD1	88	2362969CB1	
4753527	32	4753527CD1	89	4753527CB1	
6928688	33	6928688CD1	90	6928688CB1	
7506388	34	7506388CD1	91	7506388CB1	
7376372	35	7376372CD1	92	7376372CB1	
2754344	36	2754344CD1	93	2754344CB1	
8268822	37	8268822CD1	94	8268822CB1	
1814553	38	1814553CD1	95	1814553CB1	4521921CA2
71217830	39	71217830CD1	96	71217830CB1	3056930CA2
7506252	40	7506252CD1	97	7506252CB1	
2270608	41	2270608CD1	98	2270608CB1	3805161CA2
7502428	42	7502428CD1	99	7502428CB1	
368741	43	368741CD1	100	368741CB1	7667604CA2
7506379	44	7506379CD1	101	7506379CB1	
7506253	45	7506253CD1	102	7506253CB1	8736487CA2
7506353	46	7506353CD1	103	7506353CB1	
7506372	47	7506372CD1	104	7506372CB1	
7506335	48	7506335CD1	105	7506335CB1	
5546982	49	5546982CD1	106	5546982CB1	
7507432	50	7507432CD1	107	7507432CB1	
5639578	51	5639578CD1	108	5639578CB1	
7509080	52	7509080CD1	109	7509080CB1	
7505899	53	7505899CD1	110	7505899CB1	
7505904	54	7505904CD1	111	7505904CB1	
7509224	55	7509224CD1	112	7509224CB1	
7505922	56	7505922CD1	113	7505922CB1	
7507695	57	7507695CD1	114	7507695CB1	

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
1	7506140CD1	g2696611	0.0	[Rattus norvegicus] RNA splicing-related protein Imai, Y., et al. (1998) Cloning of a gene, YT521, for a novel RNA splicing-related protein induced by hypoxia/reoxygenation. Brain Res. Mol. Brain Res. 53, 33-40
		328064 Rn.2155	0.0	[Rattus norvegicus][RNA-binding protein; Small molecule-binding protein] YT521, a nuclear protein with glutamic acid, proline and arginine rich regions, predicted to be a component of the spliceosomal complex, may play a role in the determination of pre-mRNA splice site selection Imai, supra.
		661120 FLJ21940	7.0E-28	[Homo sapiens] Protein of unknown function, has a region of high similarity to a region of rat Rn.2155, which interacts with splicing factors and is induced by oxygen free radicals
2	1889415CD1	g6573115	0.0	[Mus musculus] p300 transcriptional cofactor JMY Shikama, N., et al. (1999) A novel cofactor for p300 that regulates the p53 response. Mol. Cell 4, 365-376
		618568 jmy	0.0	[Mus musculus][Transcription factor] Junction mediating and regulatory protein, a transcription cofactor that forms a coactivator complex with p300 (CBP) and augments p53 (Trp53)-dependent transcription and apoptosis Partanen, A., et al. (1999) Developmentally regulated expression of the transcriptional cofactors/histone acetyltransferases CBP and p300 during mouse embryogenesis. Int. J. Dev. Biol. 43, 487-94
3	7506047CD1	g1511630	0.0	[Homo sapiens] homeodomain protein Zinovieva, R.D., et al. (1996) Structure and chromosomal localization of the human homeobox gene Prox 1. Genomics 35, 517-522

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
3	cont	337236 PROX1	0.0	[Homo sapiens][Transcription factor; DNA-binding protein] Prospero-related homeobox 1, a homeodomain-containing transcription factor that may be involved in development of the lens Torii Ma, et al. (1999) Transcription factors Mash-1 and Prox-1 delineate early steps in differentiation of neural stem cells in the developing central nervous system. Dev Suppl 126, 443-56
		325588 Prox1	0.0	[Mus musculus][Transcription factor; DNA-binding protein] Prospero-related homeobox 1, a member of the homeodomain-containing family of transcription factors, involved in the development of the lens, the lymphatic and central nervous systems, also plays a role in apoptosis, cell migration, and proliferation Wigle, J. T., and Oliver, G. (1999) Prox1 function is required for the development of the murine lymphatic system. Cell 98, 769-78.
4	7505849CD1	g13559367	9.7E-79	[Homo sapiens] mitochondrial ribosomal protein L11 (L11mt) Suzuki, T., et al. (2001) Structural Compensation for the Deficit of rRNA with Proteins in the Mammalian Mitochondrial Ribosome. SYSTEMATIC ANALYSIS OF PROTEIN COMPONENTS OF THE LARGE RIBOSOMAL SUBUNIT FROM MAMMALIAN MITOCHONDRIA J. Biol. Chem. 276, 21724-21736
		475577 MRPL11	8.5E-80	[Homo sapiens][RNA-binding protein; Ribosomal subunit][Cytoplasmic; Mitochondrial] Protein of the large 60S ribosomal subunit, has moderate similarity to S. cerevisiae Mrp19p, which is a mitochondrial ribosomal protein of the large subunit
		276059 B0303.15	2.4E-29	[Caenorhabditis elegans][RNA-binding protein; Ribosomal subunit][Cytoplasmic] Member of the ribosomal protein L12 protein family
5	7505972CD1	g12654167	1.6E-41	[Homo sapiens] cold inducible RNA-binding protein

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
5 cont		334674 CIRBP	1.4E-42	[Homo sapiens][RNA-binding protein][Nuclear] Cold inducible RNA-binding protein, an RNA-binding protein that suppresses cell proliferation in response to cold shock Nishiyama, H., et al. (1997) Cloning and characterization of human CIRP (cold-inducible RNA-binding protein) cDNA and chromosomal assignment of the gene. Gene 204, 115-20
		584099 Cirbp	2.8E-42	[Mus musculus][RNA-binding protein][Nuclear] Cold inducible RNA-binding protein, a glycine-rich RNA-binding protein that suppresses cell proliferation in response to cold shock, may play a role in biological rhythms Nishiyama, H., et al. (1998) Diurnal change of the cold-inducible RNA-binding protein (Cirp) expression in mouse brain. Biochem Biophys Res Commun 245, 534-8
6	750599 CDI	g12061189 4755561 LOC51008	1.0E-166 9.1E-168	[Homo sapiens] ASC-1 complex subunit P50 [Homo sapiens][RNA-binding protein] Protein containing a KH domain, which binds RNA
7	7506003 CDI	g2460208	0.0	[Homo sapiens] RNA polymerase III largest subunit Sepehri, S. and Hernandez, N. (1997) The largest subunit of human RNA polymerase III is closely related to the largest subunit of yeast and trypanosome RNA polymerase III. Genome Res. 7, 1006-1019
		428328 RPC155	0.0	[Homo sapiens][Transferase; RNA polymerase subunit] Catalytic subunit of DNA directed RNA polymerase III Sepehri, supra.
		8899 RPO31	0.0	[Saccharomyces cerevisiae][Transferase; RNA polymerase subunit][Nuclear] RNA polymerase III, largest subunit
8	6483977 CDI	g12654349	3.9E-116	[Homo sapiens] splicing factor, arginine/serine-rich 7 (35kD)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
8 cont		343788[SFRS7]	3.4E-117	[Homo sapiens][Spliceosomal subunit; RNA-binding protein][Nuclear] Splicing factor arginine serine rich 7, a splicing factor of the SR protein family containing an RNA binding domain (RBD) and a serine/arginine (SR) domain that is implicated in alternative pre-mRNA splicing Cavaloc, Y., et al. (1994) Characterization and cloning of the human splicing factor 9G8: a novel 35 kDa factor of the serine/arginine protein family. Embo Journal 13, 2639-49
		588041[Sfrs3]	4.3E-39	[Mus musculus][Spliceosomal subunit; RNA-binding protein][Nuclear] Splicing factor arginine/serine-rich 3, a splicing factor of the SR family that contains an RRM (RNA recognition motif) domain, influences pre-mRNA splice site selection, regulates alternative splicing of its own mRNA, and is essential for development
9	6301777CD1	g15082360 588758[Hes2]	5.1E-36 1.6E-16	[Homo sapiens] hairy and enhancer of split (Drosophila) homolog 2 [Rattus norvegicus][Inhibitor or repressor; DNA-binding protein; Transcription factor] Hairy and enhancer of split 2, contains a basic helix-loop-helix domain, predicted to repress transcription from promoters that contain E-box and N-box sequences Satow, T., et al. (2001) The basic helix-loop-helix gene hesr2 promotes gliogenesis in mouse retina. J Neurosci 21, 1265-73
10	7505976CD1	g7673373	1.3E-60	[Homo sapiens] SCAN-related protein RAZ1 Sander, T.L., et al. (2000) Identification of a novel SCAN box-related protein that interacts with MZF1B. The leucine-rich SCAN box mediates hetero- and homoprotein associations. J. Biol. Chem. 275, 12857-12867

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
10 cont		749336 SCAND1	1.1E-61	[Homo sapiens][Small molecule-binding protein] SCAN-related protein that associates with the MZF1B alternate splice form of the zinc finger transcription factor MZF1, contains a SCAN domain, which mediates hetero- and homeoprotein associations, and an arginine-rich region Castillo, G., et al. (1999) An adipogenic cofactor bound by the differentiation domain of PPARgamma. Embo Journal 18, 3676-87
		611238 Scand1	4.6E-16	[Mus musculus][Ligand][Nuclear] Peroxisome proliferative activated receptor gamma coactivator 2, transcriptional coactivator that binds to and increases the transcriptional activity of PPAR gamma (Pparg), promotes adipogenesis
11	7506016CD1	g13278720	5.8E-273	[Homo sapiens] Similar to interferon regulatory factor 5
		743716 IRF5	5.0E-274	[Homo sapiens][DNA-binding protein; Transcription factor] Interferon regulatory factor 5, a transcription factor that induces expression of members of the interferon A family, especially IFNA8, and IFNB1 in response to NDV viral infection Pitha, P. M., et al. (1998) Role of the interferon regulatory factors (IRFs) in virus-mediated signaling and regulation of cell growth. Biochimie 80, 651-658
		429942 Irf5	4.2E-238	[Mus musculus][Transcription factor] Interferon regulatory factor 5, has strong similarity to human IRF5, which is a transcription factor that induces expression of specific members of the interferon A family and human IFNB1 in response to NDV viral infection
12	7506086CD1	g541678	0.0	[Homo sapiens] hbZ17 Luna, L., et al. (1994) Molecular cloning of a putative novel human bZIP transcription factor on chromosome 17q22. Genomics 22, 553-562

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
12 cont		336592 NFE2L1	0.0	[Homo sapiens][Activator, DNA-binding protein;Transcription factor][Nuclear] Nuclear factor erythroid derived 2-like 1, transcriptional activator, involved in heme biosynthesis, coordination of nuclear and mitochondrial gene expression, may have a role in mitochondrial genome maintenance and mitochondrial biogenesis Huo, L., and Scarpulla, R. C. (2001) Mitochondrial DNA instability and perimplantation lethality associated with targeted disruption of nuclear respiratory factor 1 in mice Mol Cell Biol 21, 644-54
		582223 Nfe2l1	7.3E-258	[Mus musculus][Activator; DNA-binding protein;Transcription factor] Nuclear factor erythroid derived 2-like 1, transcriptional activator, involved in erythropoiesis, oxidative stress response, mitochondrial genome maintenance, gastrulation, and mesoderm determination
13	4176657CD1	g6581093	2.1E-92	[Mus musculus] transposase-like protein Smit, A.F.A. (1999) Interspersed repeats and other mementos of transposable elements in mammalian genomes. Curr. Opin. Genet. Dev. 9, 657-63
14	7506056CD1	g2832260	1.0E-271	[Homo sapiens] DNA polymerase epsilon small subunit Jokela, M., et al. (1998) The small subunits of human and mouse DNA polymerase epsilon are homologous to the second largest subunit of the yeast Saccharomyces cerevisiae DNA polymerase epsilon. Nucleic Acids Res. 26, 730-734
		337072 POLE2	1.1E-270	[Homo sapiens][Regulatory subunit; Transferase; DNA polymerase or subunit] Polymerase epsilon 2, an accessory subunit of DNA polymerase epsilon, binds the catalytic subunit and stabilizes the polymerase complex. Jokela, supra
		675635 Pole2	2.0E-32	[Mus musculus][Regulatory subunit; Transferase; DNA polymerase or subunit] Polymerase epsilon 2, small subunit of DNA polymerase epsilon
15	7506185CD1	g1517816	0.0	[Homo sapiens] helicase

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
15 cont		347402 DDX12	0.0	[Homo sapiens][Hydrolase; Helicase; DNA-binding protein; ATPase][Nuclear; Nuclear nucleolus] DEAD box protein 12, a member of the DEAD/H box family of RNA or DNA helicases, expressed only in proliferating cells Amann, J., et al. (1997) Characterization of putative human homologues of the yeast chromosome transmission fidelity gene, CHL1. J Biol Chem 272, 3823-32
		248326 M03C11.2	2.7E-116	[Caenorhabditis elegans][Helicase][Nuclear] Member of the kinetochore protein (tentative) protein family
16	809661 CDI	g3328235	3.9E-92	[Xenopus laevis] 14S cohesin RAD21 subunit; member of RAD21/SCC1/MCD1 family Losada, A., et al. (1998) Identification of Xenopus SMC protein complexes required for sister chromatid cohesion. Genes Dev. 12, 1986-1997
		438159 Rad21	1.2E-90	[Mus musculus] Rad21 (S. Pombe) homolog, putative component of a cohesin complex involved in sister chromatid cohesion, cleavage during the metaphase-anaphase transition may allow chromatid separation Darwiche, N., et al. (1999) Characterization of the components of the putative mammalian sister chromatid cohesion complex. Gene 233, 39-47
		343734 RAD21	4.0E-90	[Homo sapiens] Rad21 (S. Pombe) homolog, component of a cohesin complex involved in sister chromatid cohesion, cleavage during the metaphase-anaphase transition may allow chromatid separation, expression is downregulated by hypoxia in tumor cells
17	8174603 CDI	g23273980	0.0	testis nuclear RNA-binding protein-like [Homo sapiens]
		g673456	2.2E-54	[Mus musculus] testis nuclear RNA binding protein Schumacher, J.M. et al. (1995) Biol. Reprod. 52:1274-1283.
		330218 Adarb1	2.6E-34	[Rattus norvegicus][Hydrolase; RNA-binding protein][Nuclear] Double-stranded RNA adenosine deaminase, mRNA editing of the glutamate receptor by site-specific deamination of adenosines; overexpression of human (ADARB1) may play a role in Down Syndrome phenotype development. Melcher, T. et al. (1996) J. Biol. Chem. 271:31795-31798.



Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
17 cont		334046[ADARB1]	8.2E-32	[Homo sapiens][Hydrolase; RNA-binding protein][Nuclear] Double-stranded RNA adenosine deaminase, mRNA editing of the glutamate receptor subunit B by site-specific deamination of adenosines; overexpression may play a role in Down Syndrome phenotype development. Mittaz, L. et al. (1997) Genomics 41:210-217.
18	3101042CD1	g7459861	7.7E-122	[Homo sapiens] Zinc finger protein ZNF45
		658380[ZFP93]	2.2E-125	[Homo sapiens] Member of the XRCC1-linked KRAB zinc-finger protein family, has similarity to murine Zfp93.
		609312[Zfp235]	9.6E-115	Shannon, M. et al. (1996) Genomics 33:112-120. [Mus musculus][Inhibitor or repressor; Transcription factor; DNA-binding protein] Protein containing sixteen C2H2 type zinc finger domains, which bind nucleic acids, and a KRAB (kruppel-associated box) domain which may mediate transcriptional repression.
19	4972035CD1	g14278861	0.0	[Homo sapiens] PHD zinc finger transcription factor Yochum, G. S. and Ayer, D. E. (2001) Mol. Cell. Biol. 21: 4110-4118.
		738566[PF1]	0.0	[Homo sapiens] mSin3A-interacting protein, has two plant homeodomain (PHD) zinc fingers, may link the TLE corepressor and the mSin3A-histone deacetylase complex to repress transcription. Yochum, et al. (2001) <i>supra</i> .
20	7506265CD1	g6006558	1.0E-61	ribosomal protein S18 [Homo sapiens]
		g3287678	1.5E-49	[Arabidopsis thaliana] Match to ribosomal S18 gene mRNA gb Z28701, DNA gb Z23165 from A. thaliana. ESTs gb T21121, gb Z17755, gb R64776 and gb R30430 come from this gene.
		252690[Y57G11 C.16]	3.8E-51	[Caenorhabditis elegans][RNA-binding protein][Cytoplasmic] Member of the ribosomal protein S18 protein family.

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
20 cont		430218 Rps18	1.3E-46	[Mus musculus][RNA-binding protein; Ribosomal subunit][Cytoplasmic] Ribosomal protein S18, a putative ribosomal protein; human RPS18 is overexpressed in various tumors. MacMurray, A. J., and Shin, H. S. (1992) Mamm. Genome 2: 87-95.
21	7506304CD1	g12652649	1.8E-28	[Homo sapiens] ribosomal protein L28
		659262 Rpl28	2.1E-54	[Rattus norvegicus][Structural protein;RNA-binding protein; Ribosomal subunit][Cytoplasmic] Ribosomal protein L28, component of the large 60S ribosomal subunit; human RPL28 gene shows abnormal expression in human colorectal carcinoma cells. Wool, I. G. et al. (1990) Biochim. Biophys. Acta 1050: 69-73.
		477258 Rpl28	4.0E-29	[Mus musculus][Structural protein] Ribosomal protein L28, component of the large 60S ribosomal subunit; human RPL28 gene shows abnormal expression in human colorectal carcinoma cells. Burke, P. S. et al. (1994) Gene 142:315-316.
22	7506198CD1	g11139704	0.0	[Homo sapiens] Pumilio 2
		697504 PUM2	0.0	[Homo sapiens][RNA-binding protein] Protein containing eight Pumilio-family (Puf) domains, which bind RNA, has a region of low similarity to a region of S. cerevisiae Mpt5p, which is required for high temperature growth, for recovery from alpha-factor arrest, and for normal lifespan.
		371884 SPAC222.02c	1.6E-96	[Schizosaccharomyces pombe] Pumilio family protein.
24	6803876CD1	g6526355	5.0E-63	[Homo sapiens] hMBF1alpha
		339778 EDF1	4.3E-64	Kabe, Y. et al. (1999) J. Biol. Chem. 274:34196-34202. [Homo sapiens] Endothelial differentiation-related factor 1, a putative transcriptional coactivator that binds to calmodulin (CALM1), in a calcium-dependent manner, and to the TATA-binding protein (TBP), may be involved in endothelial cell growth and differentiation. Mariotti, M. et al. (2000) J. Biol. Chem. 275:24047-24051.

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
24 cont		618910 Edf1	1.1E-63	[Mus musculus] Protein with very strong similarity to human EDF1, a putative transcriptional coactivator that binds to calmodulin (CALM1) and to the TATA-binding protein (TBP) and that may be involved in endothelial cell growth and differentiation.
25	750628 CD1	g1289371	7.5E-179	[Homo sapiens] Ikaros/LyF-1 homolog Nietfeld, W. and Meyerhans, A. (1996) Immunol. Lett. 49:139-141.
		343040 ZNFN1A1	6.5E-180	[Homo sapiens][DNA-binding protein] Zinc finger protein subfamily 1A 1 (Ikaros), a zinc finger transcription factor that regulates the lymphopoietic system development and homeostasis; alterations in corresponding gene expression contribute to hematological malignancies. Georgopoulos, K. et al. (1994) Cell 79:143-56.
		581109 Znfn1a1	1.2E-156	[Mus musculus][Hydrolase; Activator; DNA-binding protein; Transcription factor][Nuclear] Zinc finger protein subfamily 1A 1 (Ikaros), a zinc-finger transcription factor regulating the lymphopoietic system development and homeostasis, targets chromatin remodeling factors; human ZNFN1A1 expression is altered in human hematological malignancies. Taubenberger, J. K. et al. (1996) Cell. Immunol. 171:41-47.
26	7506175 CD1	g14211691	1.7E-238	[Homo sapiens] elongation factor 1A binding protein Mansilla, F. et al. (2001) Thesis: Structural Biology, IMSB Aarhus University, Gustav Wieds vej 10 C, Aarhus 8000 C, Denmark.
		731425 BPOZ	7.0E-159	[Homo sapiens] BPOZ protein, contains an ankyrin repeat, bipartite nuclear localization signal, and two BTB/POZ domains, inhibits cell proliferation, may be involved in protein complex assembly and developmental processes; decreased expression is seen in ovarian tumors. Unoki, M., and Nakamura, Y. (2001) Oncogene 20:4457-4465.
27	7506303 CD1	g12803179	1.0E-220	[Homo sapiens] splicing factor 3a, subunit 3, 60kD
28	7353336 CD1	g306864	2.2E-44	[Homo sapiens] high mobility group protein 17

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
28 cont		344416 HMG17	1.9E-45	[Homo sapiens][DNA-binding protein][Nuclear] High mobility group 17, binds DNA with low specificity, involved in regulation of transcription and cell differentiation, overexpressed in chronic myelogenous leukemia and may be mutated in various neoplasms. Kondos, H. et al. (1995) Biochem. Mol. Biol. Int. 36:803-809.
		587719 Hmg17	4.6E-44	[Mus musculus][DNA-binding protein][Nuclear] High mobility group 17, binds DNA with variable sequence specificity, involved in regulation of transcription and cell differentiation; human HMG17 is overexpressed in chronic myelogenous leukemia. Bustin, M. et al. (1995) DNA Cell. Biol. 14:997-1005.
29	3001652CD1	g12483904	1.5E-117	[Rattus norvegicus] zinc finger protein HIT-39
		310825 Hs.18375	0.0	[Homo sapiens] [Inhibitor or repressor; Transcription factor] Protein containing KRAB (kruppel-associated box) domains which may mediate protein-protein interactions, contains C2H2 type zinc finger domains, which bind nucleic acids
		308339 ZNF184	5.2E-82	[Homo sapiens] Zinc finger protein 184, a member of the Kruppel zinc finger protein family, contains tandemly repeated C2H2-type zinc finger motifs at the C-terminus, highly expressed in testis Goldwurm, S. et al. (1997) Genomics 40:486-489 Identification of a novel Krueppel-related zinc finger gene (ZNF184) mapping to 6p21.3.
30	1689128CD1	g1216477	7.9E-21	[Mus musculus] zinc finger protein 60 Perez, M. et al. (1996) FEBS Lett. 387:117-121 Zfp60, a mouse zinc finger gene expressed transiently during in vitro muscle differentiation.

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
30 cont		581073 Zfp37	1.6E-23	[Mus musculus] [DNA-binding protein; Transcription factor] Zinc finger protein-37, member of the KRAB (kruppel-associated box) zinc-finger protein family, is expressed in brain and testis and may regulate spermiogenesis de Luis, O., Lopez-Fernandez, L. A., and del Mazo, (1999) J. Exp Cell Res 249:320-326 Tex27, a gene containing a zinc-finger domain, is up-regulated during the haploid stages of spermatogenesis.
		567485 OAZ	3.3E-21	[Homo sapiens] [DNA-binding protein] Protein with very strong similarity to rat Rn.9981, which is a zinc finger protein that regulates olfactory neuronal differentiation by interacting with Olf-1/EBF transcription factors, contains twenty eight C2H2 type zinc finger domains
31	2362969CD1	g5360107	1.3E-117	[Homo sapiens] NY-REN-36 antigen Scanlan, M.J. et al. (1999) Int. J. Cancer 83:456-464 Antigens recognized by autologous antibody in patients with renal-cell carcinoma
		423317 KIAA0295	3.9E-107	[Homo sapiens] [DNA-binding protein] Protein containing a C2H2 type zinc finger domain, which bind nucleic acids, has a region of weak similarity to a region of murine Nfh (heavy subunit of neurofilament), which is an intermediate filament
		617824 NEFH	8.8E-11	[Homo sapiens] [Structural protein] [Cytoplasmic; Cytoskeletal] Heavy polypeptide of neurofilament, a structural protein of the cytoskeleton that likely regulates axonal caliber and synaptic transmission; variants may be associated with some cases of amyotrophic lateral sclerosis Robberecht, W. (2000) J Neurol 247:2-6 Genetics of amyotrophic lateral sclerosis.

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
32	4753527CD1	g4106464	6.7E-164	[Mus musculus] AE-1 binding protein AEBP2 He, G.P. et al. (1999) J. Biol. Chem. 274:14678-14684 Cloning and characterization of a novel zinc finger transcriptional repressor. A direct role of the zinc finger motif in repression.
		429622 Aebp2	5.7E-165	[Mus musculus] [Inhibitor or repressor; DNA-binding protein; Transcription factor] [Nuclear] AE-binding protein 2, a transcriptional repressor that contains three Cys2-His2 Kruppel type zinc finger motifs and binds the promoter of fatty acid-binding protein Ap2 (Fabp4), may play a role in the control of adipocyte differentiation He, G. P. et al. (1999) J Biol Chem 274:14678-14684 Cloning and characterization of a novel zinc finger transcriptional repressor. A direct role of the zinc finger motif in repression.
		716429 ZIC4	9.2E-14	[Homo sapiens] Protein containing three C2H2 type zinc finger domains, has high similarity to a region of mouse Zic1, which is a zinc finger transcriptional activator that may be involved in cerebellar differentiation and maintenance of cerebellar granule cells
33	692868CD1	g10954044	2.3E-104	[Homo sapiens] KRAB zinc finger protein ZFQR Ran, Q. et al. (2001) Exp. Cell Res. 263:156-162 Characterization of a novel zinc finger gene with increased expression in nondividing normal human cells.
		731891 MGC44010	2.0E-265	[Homo sapiens] Protein containing 8 C2H2 type zinc finger domains, which bind nucleic acids, and a kruppel-associated box (KRAB) domain, which are involved in protein-protein interactions, has moderate similarity to human RBAK, which binds retinoblastoma protein RB1

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
33 cont		622017 ZBRK1	3.3E-105	[Homo sapiens] [Inhibitor or repressor; DNA-binding protein; Transcription factor] [Nuclear] Zinc finger and BRCA1-interacting protein with a KRAB domain 1, binds BRCA1 and represses GADD45A transcription through intron 3, inhibits quiescent cells stimulated with growth factors from entering cellcycle and maintains the nondividing state of cells Zheng, L. et al. (2000) Mol Cell 6:757-768 Sequence-Specific Transcriptional Corepressor Function for BRCA1 through a Novel Zinc Finger Protein. ZBRK1.
34	7506388CD1	g532762	4.5E-165	[Homo sapiens] AF-17 Prasad, R. et al. (1994) Proc. Natl. Acad. Sci. U.S.A. 91:8107-8111 Leucine zipper dimerization motif encoded by the AF-17 gene fused to ALL-1 (MLL) in acute leukemia.
		339960 MLLT10	0.0	[Homo sapiens] [DNA-binding protein; Transcription factor] [Nuclear] Myeloid-lymphoid or mixed-lineage leukemia translocated to 10, putative transcription factor, contains a leukemia associated protein (LAP) domain and a leucine zipper; fusion of the corresponding gene to MLL or PICALM is found in acute leukemias Salmon-Nguyen, F. et al. (2000) Cancer Genet Cytogenet 122:137-140 CALM-AF10 fusion gene in leukemias: simple and inversion-associated translocation (10;11).
		585303 MLLT10	0.0	[Mus musculus] [Inhibitor or repressor; DNA-binding protein; Transcription factor] [Nuclear; Cytoplasmic] Myeloid-lymphoid or mixed-lineage leukemia translocated to 10, putative transcription factor, contains a leukemia associated protein (LAP) domain and a leucine zipper; fusion of human MLLT10 to MLL or PICALM is found in acute leukemias

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
35	7376372CD1	g1654112	0.0	[Homo sapiens] ras-responsive element binding protein Thiagalingam, A. et al. (1996) Mol. Cell. Biol. 16:5335-5345 RREB-1, a novel zinc finger protein, is involved in the differentiation response to Ras in human medullary thyroid carcinomas.
		337828 RREB1	0.0	[Homo sapiens] [Activator; DNA-binding protein; Transcription factor] [Nuclear] Ras responsive element binding protein 1, zinc-finger protein implicated in transcriptional activation response to Ras- or Raf-induced cell differentiation
		625895 ZNF197	6.9E-44	[Homo sapiens] [Transcription factor] Zinc finger protein 197, member of the zinc finger transcription factor family, contains twenty C2H2-type zinc finger motifs, high level expression is associated with thyroid papillary carcinomas Gonsky, R., et al. (1997) Nucleic Acids Res 25:3823-3831 Identification of rapid turnover transcripts overexpressed in thyroid tumors and thyroid cancer cell lines: use of a targeted differential RNA display method to select for mRNA subsets.
36	2754344CD1	g387424	4.2E-244	[Mus musculus] mdm-1 Snyder, L.C. et al. (1988) J. Biol. Chem. 263:17150-17158 A gene amplified in a transformed mouse cell line undergoes complex transcriptional processing and encodes a nuclear protein.
		606464 MDM1	9.1E-87	[Homo sapiens] [Nuclear] Protein with high similarity to murine Mdm1, which is a nuclear protein, and the gene for which is amplified in transformed cells
		585283 Mdm1	4.3E-80	[Mus musculus] [Nuclear] Nuclear protein; gene is amplified up to 30-fold in transformed murine cells and generates a variety of alternatively spliced messages Snyder, L. C. et al. (1988) J Biol Chem 263:17150-17158 A gene amplified in a transformed mouse cell line undergoes complex transcriptional processing and encodes a nuclear protein.



Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
37	8268822CD1	g5360109	1.4E-90	[Homo sapiens] NY-REN-37 antigen Scanlan, M.J. et al. (1999) Int. J. Cancer 83:456-464 Antigens recognized by autologous antibody in patients with renal-cell carcinoma.
		691420 FLJ11806	1.3E-128	[Homo sapiens] Antigen recognized by autologous antibody in a patient with renal cell carcinoma
38	1814553CD1	g514388	8.6E-13	[Musca domestica] transposase Atkinson, P.W. et al. (1993) Proc. Natl. Acad. Sci. U.S.A. 90:9693-9697 The hobo transposable element of Drosophila can be cross-mobilized in houseflies and excises like the Ac element of maize.
39	71217830CD1	g9837581	1.8E-26	[Drosophila melanogaster] WIBG Ohlstein, B. et al. (2000) Genetics 155:1809-1819 The drosophila cystoblast differentiation factor, benign gonial cell neoplasm, is related to DExH-box proteins and interacts genetically with bag-of-marbles.
40	7506252CD1	g6440969	0.0	[Homo sapiens] RECQL4 helicase Kitao, S. et al. (1999) Genomics 61 (3):268-276 Rothmund-thomson syndrome responsible gene, RECQL4: genomic structure and products.
		341164 RECQL4	0.0	[Homo sapiens] [Hydrolase; Helicase; DNA-binding protein; ATPase] RecQ protein-like 4, a member of the RecQ family of DNA helicases that may play a role in DNA repair; mutations in the corresponding gene are found in patients with Rothmund-Thomson syndrome, which is associated with genomic instability and cancer Kitao, S. et al. (1999) Nat. Genet. 22:82-84 Mutations in RECQL4 cause a subset of cases of Rothmund-Thomson syndrome.
		662679 Recql	2.2E-22	[Mus musculus] [Hydrolase; Helicase; DNA-binding protein; ATPase] [Nuclear] DNA helicase Q1, has similarity to Escherichia coli RecQ, which is an ATP-dependent DNA helicase, highly expressed in testis and thymus

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
41	2270608CD1	g13905190	1.6E-66	[Mus musculus] Similar to TATA box binding protein (TBP)-associated factor, RNA polymerase II, C1, 130kD
42	7502428CD1	g9650982	4.0E-125	[Homo sapiens] testis-specific RING Finger protein Yoshikawa, T. et al. (2000) Biochim. Biophys. Acta 1493:349-355 Isolation of a cDNA for a novel human RING finger protein gene, RNF18, by the virtual transcribed sequence (VTS) approach(I).
		610890 RNF18	3.5E-126	[Homo sapiens] Ring finger protein 18, a C3HC4 type member of the RING finger that may be a component of ribonucleoprotein complexes, expressed preferentially in testis
		743912 SSA1	1.1E-33	[Homo sapiens] [DNA-binding protein; RNA-binding protein] Sjogren syndrome antigen A1 (52 kD), a member of the Tripartite Motif protein family and component of Ro-SSA ribonucleoprotein complexes; recombined by autoantibodies in Sjogren's Syndrome, maternal antibodies are linked to neonatal lupus erythematosus Eftekhar, P. et al. (2000) Eur J Immunol 30:2782-2790 Anti-SSA/Ro52 autoantibodies blocking the cardiac 5-HT4 serotonergic receptor could explain neonatal lupus congenital heart block.
43	368741CD1	g55471	7.9E-122	[Mus musculus] Zfp-29 Denny, P. and Ashworth, A. (1991) A zinc finger protein-encoding gene expressed in the post-meiotic phase of spermatogenesis. Gene 106:221-227
		584163 Zfp29	6.7E-123	[Mus musculus][Transcription factor; DNA-binding protein] Zinc finger protein 29, a putative transcription factor that may regulate post-meiotic germ cell gene expression, expressed specifically in post-meiotic round spermatids de Luis, O. et al. (1999) Tex27, a gene containing a zinc-finger domain, is up-regulated during the haploid stages of spermatogenesis. Exp. Cell. Res. 249:320-326

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
43 cont		432838 ZNF180	1.4E-122	[Homo sapiens] Zinc finger protein 180, a putative zinc-finger protein that may play a role in developmental processes Wang, R. et al. (1996) Identification of a locus of zinc finger genes in human chromosome 19q13.1-q13.3 region by fluorescence in situ hybridization. Somat. Cell Mol. Genet. 22:245-248
44	7506379CD1	g3702270	7.3E-70	[Homo sapiens] ribosomal protein L18a
		625915 RPL18A	6.2E-71	[Homo sapiens][Structural protein; RNA-binding protein; Ribosomal subunit] [Cytoplasmic] Ribosomal protein L18a, a component of the 60S ribosomal subunit Kenmochi, N. (1998) A map of 75 human ribosomal protein genes. Genome Res. 8:509-523
		715420 E04A4.8	2.6E-41	[Caenorhabditis elegans][RNA-binding protein][Cytoplasmic] Member of the ribosomal protein L20 protein family Piano, F. (2000) RNAi analysis of genes expressed in the ovary of Caenorhabditis elegans. Curr. Biol. 10:1619-1622
45	7506253CD1	g13543433	1.4E-106	[Homo sapiens] eukaryotic translation initiation factor 4E-like 3
		339940 EIF4EL3	1.2E-107	[Homo sapiens][RNA-binding protein; Translation factor] Eukaryotic translation initiation factor 4E-like 3, a putative translation factor that binds capped mRNA Rom, E. et al. (1998) Cloning and characterization of 4EHP, a novel mammalian eIF4E-related cap-binding protein. J. Biol. Chem. 273:13104-13109
		704255  2700069E09Rik	1.5E-92	[Mus musculus][RNA-binding protein; Translation factor] [Cytoplasmic] Eukaryotic translation initiation factor 4E-like 3, a putative translation factor and capped mRNA-binding protein that may play a role in embryogenesis Temeles, G. L. et al. (1994) Expression patterns of novel genes during mouse preimplantation embryogenesis. Mol. Reprod. Dev. 37:121-129
46	7506353CD1	g9651711	2.0E-43	[Mus musculus] arsenite inducible RNA associated protein

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
46 cont		703161 Airap	1.7E-44	[Mus musculus] Arsenite inducible RNA associated protein, associates with RNA, may function in protecting cells from arsenite toxicity Sok, J. et al. (2001) Arsenite-inducible RNA-associated protein (AIRAP) protects cells from arsenite toxicity. Cell Stress Chaperones 6:6-15
		246742 aip-1	4.0E-25	[Caenorhabditis elegans] Confers C. elegans with resistance to arsenite; weakly similar to S. cerevisiae YNL155W Sok, J. et al. <i>supra</i>
47	7506372CD1	g19484175	0.0	Similar to splicing factor, arginine/serine-rich 8 (suppressor-of-white-apricot homolog, Drosophila) [Mus musculus]
		g508231	0.0	[Homo sapiens] similar to the Drosophila splicing regulator, suppressor-of-white-apricot: Swiss-Prot Accession Number P12297 Denhez, F. and Lafyatis, R. (1994) Conservation of regulated alternative splicing and identification of functional domains in vertebrate homologs to the Drosophila splicing regulator, suppressor-of-white-apricot. J. Biol. Chem. 269:16170-16179
		341234 SFRS8	0.0	[Homo sapiens][Spliceosomal subunit; RNA-binding protein] [Nuclear] Splicing factor arginine serine rich 8, a member of the SR protein family, regulates alternative splicing by influencing the selection of alternative 5' splice sites, affects alternative splicing of fibronectin, CD45 (PTPRC), and its own mRNA Sarkissian, M. et al. (1996) The mammalian homolog of suppressor-of-white-apricot regulates alternative mRNA splicing of CD45 exon 4 and fibronectin IIICS. J. Biol. Chem. 271:31106-31114
		276878 swp-1	5.2E-39	[Caenorhabditis elegans][RNA-binding protein] [Nuclear] Putative splicing regulator and member of the SWAP protein family Spikes, D. A. et al. (1994) SWAP pre-mRNA splicing regulators are a novel, ancient protein family sharing a highly conserved sequence motif with the prp21 family of constitutive splicing proteins. Nucleic Acids Res. 22:4510-4519

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
48	7506335CD1	g10121150	7.1E-80	[Homo sapiens] bHLH factor Hes4 Bessho, Y. et al. (2001) Hes7: a bHLH-type repressor gene regulated by Notch and expressed in the presomitic mesoderm. Genes Cells 6:175-185
		617990 LOC57801	6.1E-81	[Homo sapiens] Protein of unknown function, has high similarity to a region of murine Hes1 (hairly and enhancer of split), which is a helix-loop-helix negative regulator of transcription
		583195 Hes1	8.7E-34	[Mus musculus][Inhibitor or repressor; DNA-binding protein; Transcription factor] Hairy and Enhancer of split homolog 1, a helix-loop-helix negative regulator of transcription that participates in differentiation and cell fate determination of neural and other tissues, elevated levels of expression is linked to development of lymphoma
49	5546982CD1	g8648883	0.0	Cau, E. et al. (2000) Hes genes regulate sequential stages of neurogenesis in the olfactory epithelium. Dev. Suppl. 127:2323-2332
		600182 TRIM36	0.0	[Homo sapiens] zinc-binding protein
		430024 Mid2	4.9E-55	[Homo sapiens] Protein containing fibronectin type III repeats, and a C3HC4 type (RING) zinc finger, which may mediate protein-protein interactions, has low similarity to MID1, which is associated with Opitz syndrome
50	7507432CD1	g16565963	6.7E-251	[Mus musculus][Cytoplasmic; Cytoskeletal] Midline 2, a member of the B-box family of putative transcriptional regulators, contains a fibronectin type III domain and localizes to microtubules
		692064 MGC2454	5.7E-252	Buchner, G. et al. (1999) MID2, a homologue of the Opitz syndrome gene MID1: similarities in subcellular localization and differences in expression during development. Hum. Mol. Genet. 8:1397-407
				[Homo sapiens] (AF380576) SAM-dependent methyltransferase
				[Homo sapiens] Protein of unknown function, has high similarity to uncharacterized mouse 2810025A12Rik

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
50 cont		583139  Gr(ROSA)26asSo r	6.7E-29	[Mus musculus] Gene trap ROSA 26 antisense (Philippe Soriano), a ubiquitously expressed protein that is conserved in humans and in <i>C. elegans</i> Zambrowicz, B. P. et al. (1997) Disruption of overlapping transcripts in the ROSA beta geo 26 gene trap strain leads to widespread expression of beta-galactosidase in mouse embryos and hematopoietic cells. <i>Proc. Natl. Acad. Sci. U. S. A.</i> 94:3789-3794
51	5639578CD1	g3688780	2.2E-94	[Mus musculus] testis-specific Y-encoded-like protein Vogel, T. et al. (1998) Murine and human TSPYL genes: novel members of the TSPY-SET-NAPIL1 family. <i>Cytogenet. Cell Genet.</i> 81:265-270
		581749 Tspyl	1.9E-95	[Mus musculus] TSPY-like, a member of the TSPY-SET-NAPIL1 family Vogel, T. et al. <i>supra</i>
		623732 K1AA0721	9.2E-94	[Homo sapiens] Member of the NAP family of nucleosome assembly proteins
52	7509080CD1	g9857987	0.0	[Homo sapiens] SMCY Shen, P. et al. (2000) Population genetic implications from sequence variation in four Y chromosome genes. <i>Proc. Natl. Acad. Sci. U.S.A.</i> 97:7354-7359
		341282 SMCY	0.0	[Homo sapiens] Selected mouse cDNA on the Y homolog, contains a zinc finger motif and epitopes of the minor histocompatibility antigen, H-Y, a male-specific protein that elicits transplant rejection in female recipients Agulnik, A. I. et al. (1999) Mouse H-Y encoding Smcy gene and its X chromosomal homolog Smcx. <i>Mamm. Genome</i> 10:926-929
		692410 Smcx	0.0	[Mus musculus] SMC homolog X chromosome, encoded by an X-linked gene that escapes X-inactivation Agulnik, A. I. et al. (1994) A novel X gene with a widely transcribed Y-linked homologue escapes X-inactivation in mouse and human. <i>Hum. Mol. Genet.</i> 3:879-884
53	7505899CD1	g13543524	1.8E-44	[Homo sapiens] ribosomal protein, large P2

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
53 cont		337760 RPLP2	1.5E-45	[Homo sapiens][Structural protein; RNA-binding protein; Ribosomal subunit] [Cytoplasmic] Ribosomal protein large P2, an acidic phosphoprotein component of the large 60S ribosomal subunit; contains an antigen which generates autoantibodies in individuals with systemic lupus erythematosus (SLE) Takehara, K. et al. (1990) Systemic lupus erythematosus associated with antiribosomal P protein antibody. Arch. Dermatol. 126:1184-1186
		462464 p2	4.9E-24	[Aspergillus fumigatus][RNA-binding protein; Ribosomal subunit] Acidic ribosomal protein P2, acts as an allergen
54	7505904CD1	g5813803	1.1E-92	[Homo sapiens] double-stranded RNA-binding zinc finger protein JAZ Yang, M. et al. (1999) JAZ requires the double-stranded RNA-binding zinc finger motifs for nuclear localization. J. Biol. Chem. 274:27399-27406
		428812 JAZ	9.2E-94	[Homo sapiens][RNA-binding protein][Nuclear nucleolus; Nuclear] Just another zinc finger protein, contains four C2H2 zinc finger motifs, binds double-stranded RNA and RNA/DNA hybrids, induces apoptosis when overexpressed in mouse fibroblasts Yang, M. et al. <u>supra</u>
		429954 Zfp346	3.8E-81	[Mus musculus][RNA-binding protein][Nuclear] Zinc finger protein 346 (just another zinc finger protein), contains four C2H2 zinc finger motifs, binds double-stranded RNA and RNA/DNA hybrids, induces apoptosis when overexpressed in fibroblasts Yang, M. et al. <u>supra</u>
55	7509224CD1	g18307966	0.0	splicing factor, arginine/serine-rich 12 [Homo sapiens]
		g7158880	4.8E-225	[Rattus norvegicus] serine-arginine-rich splicing regulatory protein SRP86 Barnard, D.C. and Patton, J.G. (2000) Identification and characterization of a novel serine-arginine-rich splicing regulatory protein. Mol. Cell. Biol. 20:3049-3057

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
55 cont		610045 Srp86	4.1E-226	[Rattus norvegicus][Spliceosomal subunit; RNA-binding protein] [Nuclear] Serine arginine-rich splicing regulatory protein 86, contains an RNA recognition motif and serine-arginine-rich domains, interacts with other serine-arginine-rich splicing factors, involved in splicing regulation and differential splice site selection Barnard, D. C., and Patton, J. G. <i>supra</i>
		241942 D2089.1	4.5E-46	[Caenorhabditis elegans][RNA-binding protein][Nuclear] An SR protein, thought to be involved in mRNA splicing Longman, D. et al. (2000) Functional characterization of SR and SR-related genes in <i>Caenorhabditis elegans</i> . <i>Embo J.</i> 19:1625-1637
56	7505922CD1	g187291	1.7E-112	[Homo sapiens] MAD3 Haskill, S. et al. (1991) Characterization of an immediate-early gene induced in adherent monocytes which encodes ikB-like activity. <i>Cell</i> 65:1281-1289
		613187 NFKBIA	1.4E-113	[Homo sapiens][Inhibitor or repressor][Nuclear; Cytoplasmic] Nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor alpha, involved in transcription regulation by inhibiting nuclear localization and DNA binding of transcription activator NFkB in unstimulated cells; mutated in Hodgkin lymphoma Jungnickel, B. et al. (2000) Clonal deleterious mutations in the IkappaBalpha gene in the malignant cells in Hodgkin's lymphoma. <i>J. Exp. Med.</i> 191:395-402
		585391 NFKbia	2.8E-99	[Mus musculus][Inhibitor or repressor][Nuclear; Cytoplasmic] Nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor alpha, required for postnatal development and plays a role in transcription regulation by inhibiting transcriptional activator NFkB; human NFKBIA is mutated in Hodgkin lymphoma Huxford, T. et al. (1998) The crystal structure of the IkappaBalpha/NF-kappaB complex reveals mechanisms of NF-kappaB inactivation. <i>Cell</i> 95:759-770
57	7507695CD1	g15488944	6.6E-292	[Homo sapiens] Similar to zinc finger protein 35 (clone HF.10)



Table 2

Polypeptide SEQ ID NO:	Incye Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
57 cont		338990 ZNF35	4.8E-255	[Homo sapiens][Activator; Transcription factor, DNA-binding protein] Zinc finger protein 35, a transcriptional activator that may play role in cell cycle control; corresponding gene is down-regulated during terminal differentiation of leukemic myeloid cell lines and commonly deleted in several carcinomas Donti, E. et al. (1990) Localization of the human HF.10 finger gene on a chromosome region (3p21-22) frequently deleted in human cancers. Hum. Genet. 84:391-395
		58190 Zfp105	1.2E-233	[Mus musculus] Zinc finger protein 105, highly expressed in testis, may function in spermatogenesis Przyborski, S. A. et al. (1998) Differential expression of the zinc finger gene Zfp105 during spermatogenesis. Mamm. Genome 9:758-762

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	7506140CD1	709	S5 S35 S45 S49 S86 S91 S103 S152 S164 S169 S170 S181 S190 S247 S253 S258 S263 S292 S300 S310 S367 S376 S380 S409 S498 S519 S559 S570 S654 T57 T198 T240 T319 Y707	N196 N341 N568	BRAIN PD129722: P467-G655 BRAIN PD105763: M1-E193 BRAIN PD129718: R229-A293	BLAST_PRODOR
2	1889415CD1	986	S108 S115 S118 S155 S179 S209 S222 S234 S250 S322 S354 S371 S435 S512 S558 S564 S617 S696 S830 S868 S886 S887 S895 S900 S957 S972 T188 T218 T278 T311 T440 T447 T641 T683 T762 T982 Y409 Y572 Y647	N48	do EUKARYOTIC; RNA; RNP-1; DM07068 P09406 303-470: R657-R708 do REGULATORY; DM05091 S54986 1-980: D456-G831; P804-A869	BLAST_DOMO BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
3	7506047CD1	714	S50 S109 S113 S127 S176 S180 S203 S243 S255 S272 S349 S408 S424 S482 S500 S522 S626 S682 T14 T96 T241 T328 T456 T625 T665 Y235	N108 N250 N423	signal_cleavage: M1-A39	SPSCAN
					HOMEBOX DNABINDING NUCLEAR PROTEIN PROSPERO-LIKE PROX1 PROX TRANSCRIPTION REGULATION DEVELOPMENTAL PD041924: M1-K75; S67-P352	BLAST_PRODOM
					HOMEBOX PROTEIN NUCLEAR DNABINDING TRANSCRIPTION REGULATION DEVELOPMENTAL ALTERNATIVE SPLICING PROSPERO-LIKE PD017936: S353-L711	BLAST_PRODOM
					PROSPERO-LIKE DOMAIN DM06382 P29617 834- 1402: L557-L711; L325-F348; Q192-E221; S409- P514	BLAST_DOMO
					PROSPERO-LIKE DOMAIN DM06382 P34522 I-585: D541-L711	BLAST_DOMO
4	7505849CD1	166	S121 S134 T78 T87		Ribosomal protein L11: R16-V130	HMMER_PFAM
					Ribosomal protein L11 proteins BL00359: M1-E29, L40-K80, I97-V130	BLIMPS_BLOCKS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
4	cont				PROTEIN RIBOSOMAL RNABINDING L11 50S 60S L12 CHLOROPLAST L11P MITOCHONDRION PD001367: E15-V129	BLAST_PRODOM
					RIBOSOMAL PROTEIN L11 DM00681 P34264 16-161: E15-V129 DM00681 P53875 3-157: R16-V130 DM00681 P36250 1-140: E15-V130 DM00681 A64074 2-142: E15-D132	BLAST_DOMO
5	7505972CD1	142	S113 S122 S129 T43 T139 Y93		RNA recognition motif. (a.k.a. RRM, RBD, or: L8-G79	HMMER_PFAM
					Eukaryotic RNA-binding region RNP-1 proteins BL00030: L8-F26, R47-N56	BLIMPS_BLOCKS
					Eukaryotic putative RNA-binding region RNP-1 signature: F26-177	PROFILES CAN
					Nuclear transition protein 1 signature: G69-S113	PROFILES CAN
					PROTEIN RNABINDING NUCLEAR RIBONUCLEOPROTEIN REPEAT BINDING SPLICING FACTOR ALTERNATIVE HETEROGENEOUS PD000013: L8-R78	BLAST_PRODOM
					RNABINDING PROTEIN PUTATIVE COLDINDUCIBLE GLYCINERICH CIRP A18 HNRNP NUCLEAR RNPL PD050679: Y93-S136	BLAST_PRODOM
					RIBONUCLEOPROTEIN REPEAT DM00012 P98179 1-83: M1-R78 DM00012 P38159 3-84: A2-R78 DM00012 B49418 3-84: A2-R78 DM00012 S53050 1-83: M1-177	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
5	cont				Eukaryotic putative RNA-binding region RNP-1 signature: R47-F54	MOTIFS
6	7505991CD1	317	S115 S189 S233 S288 S297 T62 T80 T118 T126 Y285		KH domain: T62-G107	HMMER_PFAM
7	7506003CD1	1359	S16 S24 S63 S118 S147 S357 S364 S692 S878 S884 S893 S941 S977 S1040 S1187 T165 T181 T199 T274 T280 T338 T391 T485 T531 T553 T557 T596 T733 T743 T843 T889 T1009 T1017 T1049 T1117 T1124 T1145 T1264 Y446 Y870 Y1097	N163 N355 N634 N791 N815 N992 N1062 N1075 N1218	KH domain proteins famil PF00013: I73-I84 RNA polymerase alpha subunit: S248-G906	BLIMPS_PFAM HMMER_PFAM
					RNA polymerase A/beta'/A" subunit: I1006-A1335	HMMER_PFAM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
7 cont					Eukaryotic RNA polymerase II heptapeptide repeat proteins BL00115: H48-I86, D87-S118, H157-K178, L226-V266, G300-I328, I328-G358, K359-V386, A387-N419, G447-D501, G502-A527, N528-A569, C632-N674, A676-C724, E746-M795, I796-A835, N836-T865, A866-S892, S893-C934, P1029-N1062, I1063-I1085, H1182-S1217, N1218-M1263	BLIMPS_BLOCKS
					RNA POLYMERASE DNADIRECTED TRANSCRIPTION TRANSFERASE SUBUNIT CHAIN LARGEST PROTEIN ZINC PD000656: G327-G906; L250-N662	BLAST_PRODROM
					DNADIRECTED RNA POLYMERASE III LARGEST SUBUNIT EC 2.7.7.6 RPC155 TRANSFERASE TRANSCRIPTION ZINC ZINC FINGER NUCLEAR PROTEIN PD097339: L907-G988	BLAST_PRODROM
					RNA POLYMERASE SUBUNIT DNADIRECTED TRANSFERASE TRANSCRIPTION ZINC LARGEST PROTEIN NUCLEAR PD001238: V11-R131	BLAST_PRODROM
					DNADIRECTED RNA POLYMERASE LARGEST SUBUNIT III TRANSFERASE TRANSCRIPTION ZINC ZINC FINGER PD020898: P132-D249	BLAST_PRODROM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
7	cont				DNA-DIRECTED RNA POLYMERASE II DM00252 P04051 I01-654; I104-Q641 DM00252 P08968 I08-641; I104-V638 DM00252 P31813 S3-611; L171-D609; C69-A161 DM00252 P27625 I1-751; Q422-V638; L142-I421; C69-L250	BLAST_DOMO
8	6483977CD1	226	S89 S123 S128 S144 S159 S167 S175 S183 S208 S215 Y14		RNA recognition motif. (a.k.a. RRM, RBD, or: V13-V79  Protamine P1 proteins BL00048: S123-R149 Eukaryotic putative RNA-binding region RNP-1 signature: F31-V77	HMMER_PFAM  BLIMPS_BLOCKS PROFILES SCAN
					SPLICING FACTOR, ARGININE/SERINERICH 7 FACTOR 9G8 NUCLEAR PROTEIN RNABINDING MRNA ALTERNATIVE PHOSPHORYLATION REPEAT PD056023: E80-H120	BLAST_PRODROM
					RIBONUCLEOPROTEIN REPEAT DM00012 P23152 S-82; E10-G84 DM00012 A46398 6-83; K12-G84 DM00012 Q09167 I-72; K12-E80	BLAST_DOMO
					PAPILLOMA VIRUS E2 PROTEIN DM00171 P36783 I-378; R124-E223 signal_cleavage: M1-G51	BLAST_DOMO
9	6301777CD1	76	S16	N30	ATP synthase gamma subunit signature: D9-G45	SPSCAN PROFILES SCAN

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
9					MYC-TYPE, 'HELIX-LOOP-HELIX' DIMERIZATION DOMAIN DM00051 P35429 1-95; M1-E47	BLAST_DOMO
10	7505976CD1	124	S29 S39 T63 T106 T118		signal_cleavage: M1-P19	SPSCAN
11	7506016CD1	488	S158 S267 S291 S297 S342 S430 S436 T71 T238	N2 N95	Interferon regulatory factor transcription factor domain: T10-N123	HMMER_PFAM
					Tryptophan pentad repeat proteins (IRF family) proteins BL00601: R16-G54, D79-P107, N378-G390	BLIMPS_BLOCKS
					Tryptophan pentad repeat (IRF family) signature: T71-P127	PROFILESCAN
					Interferon regulatory factor signature PR00267: R16-W35, L42-P55, D60-G77, P83-D105	BLIMPS_PRINTS
					INTERFERON PROTEIN TRANSCRIPTION REGULATION DNABINDING NUCLEAR FACTOR ACTIVATOR REGULATORY INDUCTION IRF5 PD003535: G195-E418; N123-D162 PD119810: G195-I282 PD002355: P11-V120 PD023740: P194-M419	BLAST_PRODOM



Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
11	cont				IRF FAMILY DM02899 Q02556 1-317: R16-V120; P218-E354; D345-F360 DM02899 I49359 1-355: R16-E354 DM02899 S57837 1-355: R16-E354 DM02899 Q00978 1-319: R16-D162; S283-A341; P197-L228	BLAST_DOMO
					Tryptophan pentad repeat (IRF family) signature: W35-A68	MOTIFS
12	7506086CD1	576	S3 S143 S195 S244 S252 S278 S285 S306 S319 S332 S359 S386 S403 S416 S510 T79 T134 T168 T302 T558 Y77	N378	signal_cleavage: M1-V28	SPSCAN
					bZIP transcription factors basic domain proteins BL00036: K467-K479	BLIMPS_BLOCKS
					Fos transforming protein signature PR00042: N468-L484, L486-F507	BLIMPS_PRINTS
					FACTOR NUCLEAR RELATED DNABINDING PROTEIN ERYTHROID TRANSCRIPTION REGULATION ACTIVATOR NFE2 PD011575: M1-S252, L257-I320	BLAST_PRODROM
					FACTOR NUCLEAR PROTEIN DNABINDING TRANSCRIPTION REGULATION ERYTHROID RELATED ACTIVATOR NFE2 PD006162: Q324-I462	BLAST_PRODROM

Table 3

SEQ ID NO:	Incye Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
12	cont				BZIP TRANSCRIPTION FACTORS BASIC DOMAIN DM00107 A49672 534-692: C368-G527 DM00107 A49671 185-334: S416-F526 DM00107 I59340 392-549: H379-F526 do TRANSCRIPTION; NRFI; BASIC; ZIPPER; DM04996 A49672 161-441: P161-S224; G267-L337	BLAST_DOMO
					bZIP transcription factors basic domain signature: R463-R478, R464-R478, R465-R478	MOTIFS
13	4176657CD1	573	S116 S136 S256 S358 S391 S445 S463 S510 S546 T55 T79 T95 T183 T189 T218 T225 T485 T538 Y371	N216 N291	Cytosolic domain: T534-H573 Transmembrane domain: I514-L533 Non-cytosolic domain: M1-S513	TMHMMER
14	7506056CD1	501	S35 S37 S65 S77 S128 S242 S246 S315 S332 S482 S489 T96 T115 T149 T224 T318 T464 T492	N240 N298 N490	DNA POLYMERASE EPSILON SUBUNIT B II TRANSFERASE DNADIRECTED REPLICATION DNABINDING PD018390: K248-L498; L83-P303; K47-E80	BLAST_PRODUM
					DNA POLYMERASE EPSILON SUBUNIT B EC 2.7.7.7 II TRANSFERASE DNADIRECTED REPLICATION DNABINDING NUCLEAR PROTEIN PD144164: M1-E46	BLAST_PRODUM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
15	7506185CD1	893	S21 S96 S192 S204 S212 S291 S485 S505 S520 T5 T48 T104 T180 T471 T477 T512 T556 T594	N3 N784 N829	DEAD and DEAH box families ATP-dependent helicases signatures: L367-Q417	PROFLESCAN
					HELICASE PROTEIN ATPBINDING DNABINDING ATPDEPENDENT DNA REPAIR PROBABLE NUCLEAR DING PD004168: L563-V869; V748-S890	BLAST_PRODROM
					HELICASE PROTEIN ATPBINDING DNABINDING NUCLEAR DNA REPAIR CHL1 DNAREPAIR COMPLEMENTING PD009997: M1-D217 PD004698: K247-N464	BLAST_PRODROM
					HELICASE CHL1 PROTEIN POTENTIAL PD024600: A435-P562	BLAST_PRODROM
					DEAH-BOX SUBFAMILY ATP-DEPENDENT HELICASES DM03675[P26659]180-727: E543-R887; G344-L438 DM03675[P18074]181-728: A342-R866 DM03675[I48087]181-728: A342-R866 DM03675[P06839]182-728: K345-I865	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
16	809661ICD1	555	S9 S46 S136 S157 S169 S271 S307 S317 S383 S424 S425 S463 S471 S495 S512 T28 T88 T109 T142 T272 T331 T413 T429 T478 Y105	N134 N178 N265	PROTEIN REPAIR RAD21 DNA DOUBLESTRANDBREAK DAMAGE NUCLEAR COHESIN HOMOLOG CALCIUMBINDING PD010068: M1-K86	BLAST_PRODUM
					REPAIR PROTEIN RAD21 DNA DOUBLESTRANDBREAK HOMOLOG DAMAGE NUCLEAR CALCIUMBINDING PHOSPHORYLATION PD018225: E102-E407 PD015293: E473-I555 do PW29: CALCIUM; DM08108 JC4248 I-533: M1-D229; D148-I418 DM08108 P30776 I-544: M1-D118; D216-T321	BLAST_PRODUM
17	8174603CD1	584	S13 S41 S167 S185 S386 S396 S398 S431 S491 T329 T343 T482 T504	N489	Adenosine-deaminase (editase) domain: S396-V579, A262-R349	BLAST_DOMO HMMER_PFAM
					DEAMINASE ADENOSINE RNA PROTEIN DOUBLE-STRANDED EDITING RNA-SPECIFIC HYDROLASE ZINC RNA-BINDING PD041051: L395-K550	BLAST_PRODUM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
17	cont				DEAMINASE ADENOSINE RNA PROTEIN DOUBLE-STRANDED EDITING RNA-SPECIFIC HYDROLASE ZINC RNA-BINDING PD003961: H209-R349	BLAST_PRODOM
					DOUBLE-STRANDED RNA BINDING DOMAIN DM04852 P51400 306-710: L395-P577, L216-S364	BLAST_DOMO
					DOUBLE-STRANDED RNA BINDING DOMAIN DM04852 P55265 762-1225: H209-P577, A153-L180	BLAST_DOMO
					DOUBLE-STRANDED RNA BINDING DOMAIN DM04852 P55266 708-1174: H209-P577	BLAST_DOMO
					Leucine zipper pattern: L407-L428	MOTIFS
18	3101042CD1	554			Zinc finger, C2H2 type: Y224-H246, F196-H218, Y476-H498, Y336-H358, Y504-H526, Y364-H386, Y280-H302, Y420-H442, Y392-H414, Y448-H470, Y308-H330	HMMER_PFAM
					Zinc finger, C2H2 type, domain proteins BL00028: C478-H494	BLIMPS_BLOCKS
					C2H2-type zinc finger signature: PR00048: P279- S292, L435-G444	BLIMPS_PRINTS
					PROTEIN ZINC-FINGER METAL PD000066: H354- C366	BLIMPS_PRODOM
					PROTEIN ZINC FINGER METAL BINDING DNA BINDING ZINC FINGER PATTERNALLY EXPRESSED ZN FINGER PW1 PD017719: G304- E552, G220-F457, K244-F485, P195-H442, G164- F401	BLAST_PRODOM

Table 3

SEQ ID NO:	Incye Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
18 cont					ZINC FINGER DNA BINDING PROTEIN METAL BINDING NUCLEAR ZINC FINGER TRANSCRIPTION REGULATION REPEAT PD000072: K334-C397, R472-E536, K418-C481, K446-C509, K362-C425, K390-C453	BLAST_PRODUM
					ZINC FINGER, C2H2 TYPE, DOMAIN DM000002 Q05481 789-829: V329-E368, Q355-E396, Q411-E451	BLAST_DOMO
					ZINC FINGER, C2H2 TYPE, DOMAIN DM000002 Q05481 831-885: C313-E368, C285-E340	BLAST_DOMO
					ZINC FINGER, C2H2 TYPE, DOMAIN DM000002 P08042 314-358: C341-H386	BLAST_DOMO
					ZINC FINGER, C2H2 TYPE, DOMAIN DM000002 P52743 31-93: L323-H386	BLAST_DOMO
19	4972035CD1	1004	S40 S50 S61 S125 S134 S144 S162 S240 S241 S280 S332 S371 S376 S448 S454 S524 S529 S555 S576 S729 S837 S874 S963 S989 T54 T103 T130 T250 T257 T425 T461 T819 T862 Y857	N278 N328 N595 N606 N637 N665 N719 N856	PHD-finger: S58-R105, V273-H321	HMMER_PFAM
					Forkhead-associated (FHA) domain proteins profile BL50006: V836-A841, S858-T862	BLIMPS_BLOCKS
					PHD-finger. PF00628: C71-P85	BLIMPS_PFAM

Table 3

SEQ ID NO.	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
19					HOMEODOMAIN; PATHOGENESIS; YMR075W; DM02014 Q09819 49-172; E35-R105	BLAST_DOMO
cont					YMR075W; DM08124 Q09698 325-538; N221-V322	BLAST_DOMO
					HOMEODOMAIN; PATHOGENESIS; YMR075W; DM02014 S52835 164-321; N55-C102	BLAST_DOMO
20	7506265CD1	123	S120 T18 T110 Y66		Ribosomal protein S13/S18: V35-R113, R14-K34	HMMER_PFAM
					Ribosomal protein S13 proteins BL00646: R14-V61, D81-R113	BLIMPS_BLOCKS
					Ribosomal protein S13 signature: G73-K124	PROFILESCAN
					RIBOSOMAL PROTEIN S13 DM00495 P34788 12-148: K25-S120, I12-K34	BLAST_DOMO
					RIBOSOMAL PROTEIN S13 DM00495 P25232 12-148: I5-S120	BLAST_DOMO
					RIBOSOMAL PROTEIN S13 DM00495 P41094 12-148: K25-S120, I12-K34	BLAST_DOMO
					RIBOSOMAL PROTEIN S13 DM00495 P48151 13-149: K25-S120, I12-K34	BLAST_DOMO
					Ribosomal protein S13 signature: R92-Q105	MOTIFS
21	7506304CD1	112	S37 S67 S90 S95 T106	N12	Ribosomal L28e protein family: S2-R103	HMMER_PFAM
					PROTEIN 60S RIBOSOMAL L28 F6F22.24 PROBABLE C2E11.04 R11D1.8 PD010767: K47-P98	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
22	7506198CD1	987	S30 S82 S93 S102 S136 S182 S217 S580 S626 S891 T21 T673 T951 T956	N367 N531 N578 N619	Pumilio-family RNA binding repeat: E829-V863, E757-I791, D647-F681, G908-M942, E683-A717, E865-I899, R719-M754, A793-L827	HMMER_PFAM
					Pumilio-family RNA bindi PF00806: L648-S656, D768-K777, D840-Q848	BLIMPS_PFAM
					KIAA0099 PROTEIN PD125779: N193-F358	BLAST_PRODOR
					KIAA0099 PROTEIN PD134496: A475-E683	BLAST_PRODOR
					KIAA0099 PROTEIN PD062403: V385-I471	BLAST_PRODOR
					PROTEIN MATERNAL PUMILIO DEVELOPMENTAL REPEAT KIAA0099 PD039905: W45-P192	BLAST_PRODOR
					PUMILIO; I60K; SEGMENT; YLL013C; DM00365 P25822 1269-1339: G807-F878	BLAST_DOMO
					PUMILIO; I60K; SEGMENT; YLL013C; DM00365 A4622 1270-1339: C808-F878	BLAST_DOMO
					PUMILIO; I60K; SEGMENT; YLL013C; DM00365 P25822 1197-1267: G733-Y806	BLAST_DOMO
					PUMILIO; I60K; SEGMENT; YLL013C; DM00365 A4622 1341-1415: A879-Y957	BLAST_DOMO



Table 3

SEQ ID NO.	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
23	1381261CD1	1013	S145 S196 S261 S323 S328 S332 S397 S619 S629 S636 S669 S706 S754 S762 S827 S846 S853 S893 T353 T411 T552 T553 T571 T605 T722 T737 T836 T1007	N937	BTB/POZ domain: P14-T189	HMMER_PFAM
					Zinc finger, C2H2 type: Y309-H331, H726-H748, F765-H787, L234-C256, Y365-H388, Y337-H359	HMMER_PFAM
					Zinc finger, C2H2 type, domain proteins BL00028: C236-H252	BLIMPS_BLOCKS
					BTB (also known as BR-C) PF00651: A43-F55	BLIMPS_PFAM
					PROTEIN ZINC-FINGER METAL PD00066: H355-C367	BLIMPS_PRODOM
					POZ DOMAIN DM00509 S59069 I-171: P10-R56, A133-A161, A777-E817	BLAST_DOMO
					POZ DOMAIN DM00509 S41647 I1-189: D9-E57, V125-L150,	BLAST_DOMO
					Zinc finger, C2H2 type, domain: C311-H331, C339-H359, C367-H388, C728-H748, C767-H787	MOTIFS
					Helix-turn-helix: I81-A135	HMMER_PFAM
24	6803876CD1	141	S4 S41 S111 S134 T40 T58 T91		MULTIPROTEIN BRIDGING FACTOR EDF1 PROTEIN mRNA UNKNOWN PD019913: S4-V80	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
24 cont					REPRESSOR; CONTROL; RESTRICTION; REGULATOR; DM00461 P14327 35-98; V77-V137	BLAST_DOMO
25	7506281CD1	334	S13 S46 S47 S49 S73 S91 S92 S141 S179 S183 S199 S208 S217 S242 T23 T106 T173 T213 T269 T325 Y108	N60 N177	Zinc finger, C2H2 type, domain proteins BL00028; C279-H295	BLIMPS_BLOCKS
					PROTEIN TRANSCRIPTION REGULATION ACTIVATOR ZINC FINGER META-BINDING DNA-BINDING NUCLEAR ALTERNATIVE SPLICING PD010982; V54-D182	BLAST_PRODOM
					PROTEIN TRANSCRIPTION REGULATION ACTIVATOR ZINC FINGER METAL-BINDING DNA-BINDING NUCLEAR ALTERNATIVE SPLICING PD013352; S183-K275	BLAST_PRODOM
					PROTEIN DNA-BINDING ZINC FINGER METAL-BINDING NUCLEAR HUNCHBACK DEVELOPMENTAL GAP REPEAT TRANSCRIPTION PD004213; V276-H329	BLAST_PRODOM
					ZINC-FINGERS II DOMAIN DM03488 A56229 259-387; R201-R330	BLAST_DOMO
					ZINC-FINGERS II DOMAIN DM03488 Q03267 302-426; S208-R330	BLAST_DOMO
					IKAROS; LYF-1; TRANSCRIPTION; DNA; DM08434 A56229 197-257; G138-E200	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
25					IKAROS; LYF-1; TRANSCRIPTION; DNA; DM08434 Q03267 233-289; G138-K195	BLAST_DOMO
					Zinc finger, C2H2 type, domain: C279-H299	MOTIFS
26	7506175CD1	439	S9 S94 S231 S315 S416		signal_cleavage: M1-A64	SPSCAN
					BTB/POZ domain: E220-L337, D99-R215	HMMER_PFAM
					Ank repeat: W35-F67, M1-K34	HMMER_PFAM
					BTB (also known as BR-C) PF00651: V128-F140	BLIMPS_PFAM
					Ankyrin repeat proteins PF00651: L40-L55	BLIMPS_PFAM
					Repeat protein ankyrin nucleic acid PD00078	BLIMPS_PRODOM
					IKAROS LYF-1 TRANSCRIPTION DNA	BLAST_DOMO
					DM06485 Q10225 44-522: D5-I74, H247-Y415	
					IKAROS LYF-1 TRANSCRIPTION DNA	BLAST_DOMO
					DM06485 P40560 I-512: D5-V176, C246-V408	
27	7506303CD1	448	S102 S112 S197 S198 S234 S312 S314 S316 S415 T3 T28 T32 T78 T144 T224 T257 T261 T299 T362 T397 T422 T438 Y107 Y132	N170 N360 N395	PROTEIN SPLICESOME-ASSOCIATED NOISETTE GENE SAP SPLICING FACTOR SF3A60 T13H5.4 PD014231: Y11-G300	BLAST_PRODOM
					PROTEIN SPLICESOME-ASSOCIATED SPLICING FACTOR NOISETTE GENE SAP SF3A60 T13H5.4 PRE-MRNA PD010787: N321-L448	BLAST_PRODOM
28	7353336CD1	104	S25 S29		HMG14 and HMG17: P2-K90	HMMER_PFAM
					HMG14 and HMG17 proteins BL00355: K18-K48	BLIMPS_BLOCKS
					HMG14 and HMG17 signature: P2-P53	PROFILESCAN

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
28	cont				Nonhistone chromosomal protein HMG17 family signature PR00925: K18-P32, P34-A46, K51-D61, A70-A80	BLIMPS_PRINTS
					PROTEIN CHROMOSOMAL NONHISTONE NUCLEAR DNA-BINDING HMG14 MULTIGENE FAMILY POLYMORPHISM PD008914: P2-K104	BLAST_PRODOR
					NONHISTONE CHROMOSOMAL PROTEIN HMG-17 DM07897 P05204 1-88: P2-K90	BLAST_DOMO
					NONHISTONE CHROMOSOMAL PROTEIN HMG-17 DM04876 P12902 1-103: P2-D102	BLAST_DOMO
					NONHISTONE CHROMOSOMAL PROTEIN HMG-17 DM04876 P02316 1-99: P2-D92	BLAST_DOMO
					NONHISTONE CHROMOSOMAL PROTEIN HMG-17 DM04876 P12274 1-101: P2-A86	BLAST_DOMO
					HMG14 and HMG17 signature: R23-P32	MOTIFS
29	3001652CD1	769	S30 S39 S79 S173 S210 S285 S290 S329 S364 S366 S425 S436 S520 S529 S569 S693 T167 T279 T322 T353 T686 Y284	N180	KRAB box: V519-A581, Y29-E91	HMMER_PFAM
					Zinc finger, C2H2 type: Y312-H334, Y228-H250, H172-H194, H676-H698, Y256-H278, Y284-H306, F340-H362, F704-H726, F200-H222	HMMER_PFAM
					Zinc finger, C2H2 type, domain proteins BL00028: C230-H246	BLIMPS_BLOCKS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
29					C2H2-type zinc finger signature PR00048: P339-K352, L299-G308	BLIMPS_PRINTS
cont					PROTEIN ZINC-FINGER Metal-binding PD00066: H274-C286	BLIMPS_PRODOM
					PROTEIN ZINC FINGER ZINC PD01066: F521-G559	BLIMPS_PRODOM
					HYPOTHETICAL 87.8 KD PROTEIN HYPOTHETICAL PROTEIN ZINC-FINGER METAL-BINDING DNA-BINDING: PD175969: R363-T518; PD172267: A581-R675; PD170284: A92-L170; PD169183: R702-G769, R338-E403 KRAB BOX DOMAIN DM00605[P52738]3-77: G517-W578, G27-W88	BLAST_PRODOM
					KRAB BOX DOMAIN DM00605[Q0548]10-83: G517-S579, G27-S89	BLAST_DOMO
					KRAB BOX DOMAIN DM00605[P17097]1-76: V29-G97, V519-E589	BLAST_DOMO
					KRAB BOX DOMAIN DM00605[I48689]11-85: T518-W578, A28-Q93	BLAST_DOMO
					Cell attachment sequence: R596-D598, R749-D751	MOTIFS
					Zinc finger, C2H2 type, domain: C174-H194, C202-H222, C230-H250, C258-H278, C286-H306, C314-H334, C342-H362, C678-H698, C706-H726	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
30	1689128CD1	1081	S17 S27 S44 S55 S107 S115 S131 S186 S257 S261 S320 S364 S366 S370 S387 S442 S447 S570 S648 S663 S728 S740 S792 S809 S868 S980 S987 T39 T65 T229 T234 T754 T767 T796 T874 T932 T938 T953 T959 Y556	N35 N153 N199 N546 N748	Zinc finger, C2H2 type: F817-H839, H402-H424, F374-H396, F1019-H1041, Y897-H919, Y845-H867, F1047-H1069, F433-H455, I929-H952	HMMER_PFAM
					Zinc finger, C2H2 type, domain proteins BL00028: C819-H835	BLIMPS_BLOCKS
					ACROSIN DM0363 P12978 I0-163: P311-Q344.L310-P343, S241-P274	BLAST_DOMO
					Cytochrome c family heme-binding site signature: C1049-K1054	MOTIFS
					Immunoglobulins and major histocompatibility complex proteins signature: F1047-H1053	MOTIFS
					Zinc finger, C2H2 type, domain: C376-H396, C404-H424, C819-H839, C1021-H1041, C1049-H1069	MOTIFS

Table 3

SEQ ID NO:	Incye Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
31	2362969CD1	1007	S7 S8 S91 S112 S155 S182 S229 S302 S324 S373 S397 S448 S459 S465 S529 S544 S653 S692 S712 S716 S765 S785 S800 S886 T36 T131 T216 T220 T274 T383 T563 T617 T675 T720 T725 T891 T904 Y382 Y604	N6 N173 N223 N660 N670	Zinc finger, C2H2 type, domain proteins BL00028: H52-H68	BLIMPS_BLOCKS
					Cell attachment sequence: R619-D621	MOTIFS
					Zinc finger, C2H2 type, domain: C50-H73	MOTIFS
32	4753527CD1	511	S120 S132 S133 S148 S183 S206 S211 S217 S238 S255 S274 S358 S390 S442 S465 T81 T228	N336 N357 N414 N463	signal_cleavage: M1-G65	SPSCAN
					Zinc finger, C2H2 type: Y261-H286, F328-H352, F295-H322	HMMER_PFAM
					PROTEIN ZINC-FINGER METAL BINDING PD00066: H318-C330	BLIMPS_PRODOM
					PROTEIN NUCLEAR NUCLEOPORIN TRANSPORT PORE REPEAT COMPLEX GLYCOPROTEIN NUP214 P62 PD005717: S110- A260	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
32					ACIDIC SERINE CLUSTER REPEAT DM03496 P32583 57-405: A33-Q249 Zinc finger, C2H2 type, domain: C263-H286, C264-H286, C330-H352	BLAST_DOMO  MOTIFS
33	6928688CD1	485	S9 S17 S24 S43 S100 S140 S168 S224 S252 S280 S329 S355 S391 S429 T33 T67 T179	N3 N427	KRAB box: L23-E85  Zinc finger, C2H2 type: H214-H236, Y298-H320, Y326-H348, Y242-H264, Y270-H292, Y354-H376, H186-H208, H158-E180 Zinc finger, C2H2 type, domain proteins BL00028: C244-H260 C2H2-type zinc finger signature PR00048: P213-K226, L313-G322 PROTEIN ZINC-FINGER METAL BINDING PD00066: H232-C244 PROTEIN ZINC FINGER ZINC PD01066: F25-G63  ZINC-FINGER METAL-BINDING DNA-BINDING PATERNALLY EXPRESSED ZNFINGER PW1 PD017719: G154-I377, L145-H348, G238-R378, G164-H376, P138-H292 ZINC-FINGER METAL-BINDING DNA-BINDING NUCLEAR REPEAT TRANSCRIPTION REGULATION PD001562: L23-E85	HMMER_PFAM  HMMER_PFAM  BLIMPS_BLOCKS BLIMPS_PRINTS BLIMPS_PRODOM BLIMPS_PRODOM  BLAST_PRODOM BLAST_PRODOM



Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
33	cont				ZINC-FINGER DNA-BINDING PROTEIN METAL-BINDING NUCLEAR TRANSCRIPTION REGULATION REPEAT PD000072: K212-C275, K184-C247, K240-C303, K156-C219, K296-C359, K268-C331	BLAST_PRODUM
					MYELOBLAST KIAA0211 ZINC-FINGER METAL-BINDING DNA-BINDING PD149061: Q149-H376	BLAST_PRODUM
					KRAB BOX DOMAIN DM00605 P51523 5-79: S22-P94	BLAST_DOMO
					KRAB BOX DOMAIN DM00605 I48689 11-85: S22-P94	BLAST_DOMO
					KRAB BOX DOMAIN DM00605 P51786 24-86: E20-E79	BLAST_DOMO
					KRAB BOX DOMAIN DM00605 P52736 1-72: L23-P94	BLAST_DOMO
					ATP/GTP-binding site motif A (P-loop): G187-T194	MOTIFS
					Zinc finger, C2H2 type, domain: C188-H208, C216-H236, C244-H264, C272-H292, C300-H320, C328-H348, C356-H376	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
34	7506388CD1	I011	S4 S9 S17 S118 S204 S208 S215 S278 S303 S321 S359 S370 S376 S436 S456 S680 S686 S704 S811 T263 T286 T378 T462 T556 T663 T736 T765	N116 N134 N209 N272 N384 N442 N489 N525 N645 N648 N661 N703 N763 N810 N941	PHD-finger: G24-Q74	HMMER_PFAM
					PHD-finger: PF00628: K159-F173	BLIMPS_PFAM
					AF10 PROTEIN NUCLEAR CHROMOSOMAL TRANSLOCATION PROTOONCOGENE ZINC-FINGER MAF10 PD041462: P249-V581; PD041062: T849-D962; PD024590: H595-D833; PD132318: Q963-K1011	BLAST_PRODOR
					LEUCINE-ZIPPER DOMAIN DM07839 P55197 286-1026: T286-K1011	BLAST_DOMO
					YPR031W (includes peregrin, AF-10, AF-17) DM03695 P55197 18-284: M18-T285 DM03695 P55198 1-286: M18-S208, Q214-S274, K319-S359, G441-S513, E296-R330	BLAST_DOMO
					LEUCINE-ZIPPER DOMAIN DM07839 P55198 288-986: K305-R988	BLAST_DOMO
					Leucine zipper pattern: L750-L771, L757-L778	MOTIFS

Table 3

SEQ ID NO:	Incye Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
35	7376372CD1	1675	S110 S169 S170 S175 S180 S290 S413 S433 S463 S600 S740 S842 S883 S903 S1122 S1140 S1150 S1174 S1219 S1225 S1238 S1271 S1279 S1284 S1335 S1339 S1373 S1400 S1444 S1458 S1474 S1478 S1485 S1488 S1498 S1504 S1522 S1574 S1584 S1597 S1616 S1629 T59 T225 T229 T243 T246 T395 T514 T544 T594 T619 T709 T726 T952 T1099 T1134 T1149 T1190 T1199 T1204 T1267 T1289 T1324 T1350 T1354 T1358 T1378 T1422 T1549 T1649	N916 N1097 N1586 N1606	Zinc finger, C2H2 type: Y643-H665, L1392-H1414, Y125-H147, I1246-H1268, H97-H119, Y1539-H1561, Y671-H693, Y66-H88, K1511-H1533, Y699-H722, F316-H338, F208-H230, F790-H813, T753-H775, L235-H258	HMMER_PFAM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
35					Zinc finger, C2H2 type, domain proteins BL00028: C1394-H1410	BLIMPS_BLOCKS
cont					C2H2-type zinc finger signature PRO0048: L1526-G1535, P124-N137	BLIMPS_PRINTS
					PROTEIN ZINC-FINGER META PD00066: H1529-C1541	BLIMPS_PRODROM
					TRANSCRIPTION FACTOR RREB1 ZINC-FINGER METAL-BINDING DNA-BINDING PD184063: Q260-S669; PD041346: M932-I1246; PD050228: R697-V754; PD058224: M20-D96	BLAST_PRODROM
					ATP/GTP-binding site motif A (P-loop): A1320-S1327, A1393-S1400	MOTIFS
					Zinc finger, C2H2 type, domain: C68-H88, C99-H119, C127-H147, C210-H230, C237-H258, C318-H338, C645-H665, C701-H722, C755-H775, C792-H813, C1248-H1268, C1394-H1414, C1513-H1533, C1541-H1561	MOTIFS
36	2754344CD1	724	S54 S126 S132 S155 S268 S408 S502 S517 S577 S586 S652 S658 S696 S703 T91 T199 T294 T387 T418 T436 T459 T484 T525 T537 T569 T576 T582	N27 N215 N434 N458	TRANSFORMED MOUSE 3T3 CELL DOUBLE MINUTE 1 NUCLEAR PROTEIN MDM1 NUCLEAR PROTEIN PD142497: F213-N667, F684-K723	BLAST_PRODROM
					NUCLEAR PROTEIN TRANSFORMED MOUSE 3T3 CELL DOUBLE MINUTE MDM1 MDM1A PD037805: M1-V212	BLAST_PRODROM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
36					MDM-1; NUCLEAR; DM05421 A31794 1-222: M1-Q218	BLAST_DOMO
cont					MDM-1; NUCLEAR; DM05421 C31794 1-222: M1-V212	BLAST_DOMO
37	8268822CD1	605	S8 S82 S97 S103 S121 S129 S134 S195 S207 S268 S281 S285 S343 S390 S450 S527 S534 T59 T146 T158 T200 T257 T389 T405 T417 T517 T577 Y25 Y472 Y484 Y565	N57 N98 N101 N279 N371	Cell attachment sequence: R104-D106	MOTIFS
38	1814553CD1	865	S212 S219 S232 S278 S399 S421 S560 S568 S595 S607 S750 S850 T87 T207 T258 T409 T683 T808	N167 N230 N293 N298 N336 N359 N360 N662	Zinc finger, C2H2 type: Y59-H81, Y100-H122, Y168-H190, Y303-H325	HMMER_PFAM
					Zinc finger, C2H2 type, domain proteins BL00028: C102-H118	BLIMPS_BLOCKS
					Zinc finger, C2H2 type, domain: C61-H81, C102-H122, C170-H190, C305-H325	MOTIFS
39	71217830CD1	219	S70 S89 S132 T34 T40 T101 T163		signal_cleavage: M1-T34	SPSCAN

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
40	7506252CD1	1144	S27 S101 S134 S150 S180 S246 S257 S421 S520 S535 S798 S919 S938 S975 S1033 S1084 S1092 T47 T52 T139 T398 T465 T558 T645 T648 T713 T739 T960 T980		DEAD/DEAH box helicase: Q471-V541, I604-S679	HMMER_PFAM
					ATP-DEPENDENT DNA HELICASE, PUTATIVE PD183838: L499-A583, E563-R710	BLAST_PRODOME
					DEPENDENT; HELICASE; ATP; DNA; DM01766[P15043]16-384: E474-T722, I704-E763	BLAST_DOMO
					DEPENDENT; HELICASE; ATP; DNA; DM01766[P50729]8-376: G477-T722	BLAST_DOMO
					DEPENDENT; HELICASE; ATP; DNA; DM01766[S62467]511-893: Q479-E759	BLAST_DOMO
					DEPENDENT; HELICASE; ATP; DNA; DM01766[Q0981]511-893: Q479-E759	BLAST_DOMO
41	2270608CD1	217	S71 S101 S136 S214 T17 T196		Signal Peptide: M24-G49	HMMER
					Signal cleavage: M24-A80	SPSCAN
42	7502428CD1	309	S3 S30 S195 S278 T19 T179	N70 N293	SPRY domain: S195-H308	HMMER_PFAM
					RFP TRANSFORMING PROTEIN DM02346[P19474]59-337: E10-F193	BLAST_DOMO
					RFP TRANSFORMING PROTEIN DM01944[P18892]355-477: S195-C305	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
42					RFP TRANSFORMING PROTEIN DM01944 P19474 339-465: S195-F304	BLAST_DOMO
cont					Serpins signature: V280-I290	MOTIFS
43	368741CD1	483	S6 S43 S68 S94 S217 S348 S467 T70 T92 T97 T180 T196	N161 N413 N465	Zinc finger, C2H2 type: Y179-H201, P151-H173, F459-H481, Y291-H313, Y235-H257, Y431-H453, Y207-H229, F347-H369, F375-H397, F319-H341, Y403-H425, Y263-H285	HMMER_PFAM
					Zinc finger, C2H2 type, domain proteins BL00028: C349-H365	BLIMPS_BLOCKS
					C2H2-type zinc finger signature PR00048: P234- K247, L222-G231	BLIMPS_PRINTS
					PROTEIN ZINC-FINGER METAL BINDING PD00066: H197-C209	BLIMPS_PRODOM
					PROTEIN ZINC FINGER METAL BINDING DNA BINDING ZINC FINGER PATERNALLY EXPRESSED ZNFINGER PW1 PD017719: G175- F412, G148-H397, G343-H481	BLAST_PRODOM
					ZINC FINGER DNA BINDING PROTEIN METAL BINDING NUCLEAR ZINC FINGER TRANSCRIPTION REGULATION REPEAT PD000072: K205-C268, R373-C436, R177-C240, K289-C352, K261-C324	BLAST_PRODOM
					MYELOBLAST KIAA0211 ZINC FINGER METAL BINDING DNA BINDING PD149061: C181-P402, C265-H453	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
43	cont				ZINC FINGER, C2H2 TYPE, DOMAIN DM00002  Q05481 831-885: C380-E435, C156-P206, C408-P458  Q05481 789-829: Q366-E407, R198-E239, Q282-E323, Q394-E435, R171-K210  P08042 272-312: Q226-E267, Q394-E435  P08042 314-358: C184-H229, C156-H201, C408-H453, C212-H257, C380-H425	BLAST_DOMO
					Zinc finger, C2H2 type, domain: C153-H173, C181-H201, C209-H229, C237-H257, C265-H285, C293-H313, C321-H341, C349-H369, C377-H397, C405-H425, C433-H453, C461-H481 Ribosomal L18ae protein family: T6-F137	MOTIFS  HMMER_PFAM
44	7506379CD1	I37	S18 S19 T6 T129 T130 Y24		PROTEIN RIBOSOMAL L18A 60S L20 MULTIGENE FAMILY WUGSC: H_DJ1107K12.2 PUTATIVE YL17 PD008313:K11-V103	BLAST_PRODOM
					60S RIBOSOMAL PROTEIN L18A PD021789: K104-F137	BLAST_PRODOM
					RAT RIBOSOMAL PROTEIN L18A DM03166  S47353 1-140: Q27-F137  P41093 35-176: K11-F137  P47913 36-173: K11-F137  P05732 30-174: L12-F137	BLAST_DOMO
45	7506253CD1	200	S41 S42 S127 S175 T24 T33 T141 T148 T154 T165 T171	N22 N38 N139	Eukaryotic initiation factor 4E: Q18-L192	HMMER_PFAM



Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
45	cont				Eukaryotic initiation factor 4E proteins BL00813: K43-N86, S100-F115, M116-L159	BLIMPS_BLOCKS
					INITIATION FACTOR TRANSLATION EUKARYOTIC PROTEIN CAP-BINDING SUBUNIT 4E EIF4E EIF4E PD003697: K32-K177	BLAST_PRODROM
					PROTEIN INITIATION FACTOR CAP-BINDING 4EHP EIF4E-LIKE 4ELP TRANSLATION 4E PD101625: G14-K45	BLAST_PRODROM
					EUKARYOTIC INITIATION FACTOR 4E DM02130 [P48598]52-258: D27-G180 [P48600]1-205: V46-G183 [P06730]8-216: E29-L188 [I49644]33-217: E29-L188	BLAST_DOMO
					Eukaryotic initiation factor 4E signature: D67-W90	MOTIFS
46	7506353CD1	123	S11 S102 T14 T110		DksA/TraR zinc finger signature PR00618: C57-K68	BLIMPS_PRINTS
47	7506372CD1	874	S31 S84 S100 S137 S198 S238 S279 S283 S304 S353 S477 S519 S528 S562 S635 S665 S688 S694 S700 S731 S761 S769 S779 S816 S823 T164 T205 T425 T565 T616 T721 T765 Y265	N271	Surp module: H209-E262, Q457-Q508	HMMER_PFAM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
47					Protamine P1 proteins BL00048: R720-R746	BLIMPS_BLOCKS
cont					SPLICING FACTOR, ARGININE/SERINE RICH 8 SUPPRESSOR OF WHITE APRICOT PROTEIN HOMOLOG TRANSCRIPTION REGULATION RNA BINDING mRNA REPEAT NUCLEAR PD137630: T266-V408	BLAST_PRODROM
					SUPPRESSOR OF WHITE APRICOT PROTEIN TRANSCRIPTION REGULATION RNA BINDING mRNA SPLICING PD011756: R73-Y265	BLAST_PRODROM
					SPLICING FACTOR, ARGININE/SERINE RICH 8 SUPPRESSOR OF WHITE APRICOT PROTEIN HOMOLOG TRANSCRIPTION REGULATION RNA BINDING mRNA REPEAT NUCLEAR PD128724: N619-R673	BLAST_PRODROM
					SUPPRESSOR OF WHITE APRICOT PROTEIN HOMOLOG TRANSCRIPTION REGULATION RNA BINDING MRNA PD024346: I449-G537	BLAST_PRODROM
48	7506335CD1	189	S30 S37 S38 S41 S58 S121 T52		Myc-type, 'helix-loop-helix' dimerization domain proteins BL00038: H40-R60	BLIMPS_BLOCKS
					Myc-type, 'helix-loop-helix' dimerization domain signature: S30-A80	PROFILESKAN
					PROTEIN DNA BINDING NUCLEAR REPRESSOR TRANSCRIPTION REGULATION HAIRY FACTOR HES1 DEVELOPMENTAL PD012655: R61-P124	BLAST_PRODROM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
48 cont					MYC-TYPE, 'HELIX-LOOP-HELIX' DIMERIZATION DOMAIN DM00051  P35428 31-119: R35-E88  S29712 1-83: A32-E8  A46231 31-125: A32-E88  P14003 28-116: A32-E88	BLAST_DOMO
					Myc-type, 'helix-loop-helix' dimerization domain signature: E44-L59	MOTIFS
49	5546982CD1	716	S55 S68 S93 S155 S158 S304 S311 S312 S369 S372 S380 S403 S450 S582 S591 S597 S604 S608 S639 T11 T51 T95 T128 T179 T205 T243 T289 T322 T355 T379 T388 T459 T546 T706 Y254	N65 N123 N387 N468	B-box zinc finger.: P195-M237, A140-I180	HMMER_PFAM
					Zinc finger, C3HC4 type (RING finger): C21-L50 PROTEIN FINGER MIDLINE ZINC FINGER RING PUTATIVE TRANSCRIPTION FACTOR XPRF FETAL PD012462: N123-H232	HMMER_PFAM BLAST_PRODOM
					MIDLINE PROTEIN PUTATIVE TRANSCRIPTION FACTOR XPRF FETAL KIDNEY ISOFORM RING PD022167: R233-S403	BLAST_PRODOM
					Zinc finger, C3HC4 type (RING finger), signature: C36-V45	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
50	7507432CD1	503	S10 S47 S126 S181 S210 S292 S357 S431 T41 T54 T178 T179 T206 T222 T417 T456 Y325		Putative RNA methylase family UPF0020: R274-D434	HMMER_PFAM
51	5639578CD1	410	S2 S110 S137 S358 S362 T278		Nucleosome assembly protein (NAP): I248-Y313, S349-H388, L202-I245	HMMER_PFAM
					Nucleosome assembly protein (NAP) PF00956: I84-P94, D243-K283, F296-Y313, C348-S358	BLIMPS_PFAM
					PROTEIN NUCLEOSOME ASSEMBLY NUCLEAR ILIKE SET NAPI TESTIS SPECIFIC ASPARTIC ACID RICH PD003095: E211-S317, D356-E385	BLAST_PRODOM
					PROTEIN TESTIS SPECIFIC Y-ENCODED Y TSPY Y-ENCODED-LIKE PD035838: S323-Y381	BLAST_PRODOM
					NUCLEAR LOCALIZATION SIGNAL DOMAIN DM01696  Q01534 46-240: P191-Q367  Q01105 1-213: L205-W374  P53997 1-218: L202-W374  P13825 1-193: I208-W374	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
52	7509080CD1	1056	S20 S99 S153 S264 S329 S336 S363 S364 S396 S492 S576 S675 S723 S753 S833 S1010 S1024 T66 T166 T207 T213 T623 T685 T779 T908 T1027	N223 N248	ARID/BRIGHT DNA binding domain: E76-K172	HMMER_PFAM
					PHD-finger: I259-A307	HMMER_PFAM
					jinjC domain: W434-F550	HMMER_PFAM
					jinjN domain: E13-E59	HMMER_PFAM
					C5HC2 zinc finger: C640-L693	HMMER_PFAM
					PHD-finger PF00628: C275-P289	BLIMPS_PFAM
					PROTEIN ZK593.4 MYELOBLAST KIAA0234 SMCY XE169 SMCX ALTERNATIVE SPLICING RETINOBLASTOMA PD014548: A651-V1031	BLAST_PRODROM
					PROTEIN INTERGENIC REGION XE169 NUCLEAR ZINC FINGER ZINC FINGER METAL BINDING DNA BINDING PD005470: F316-L618, D592-L649	BLAST_PRODROM
					MYELOBLAST KIAA0234 SMCY XE169 PROTEIN SMCX ALTERNATIVE SPLICING PD022457: Q118-Y258	BLAST_PRODROM
					PROTEIN XE169 JUMONJI DEVELOPMENTAL ZK593.4 MYELOBLAST KIAA0234 SMCY SMCX ALTERNATIVE PD006843: M1-L75	BLAST_PRODROM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
52					SMCX_HUMAN XE169 DM08128[P41229]671-1190: F604-V1031	BLAST_DOMO
					FINGER; SMCX; SMCY; YDR096W; DM01930	BLAST_DOMO
					[P41229]377-669: K310-M603	
					[S44139]245-535: K310-K601	
					[P29375]346-638: Q311-M603	
53	7505899CD1	103	S19 S93		signal_cleavage: M1-A68	SPSCAN
					Signal Peptide: M1-S17	HMMER
					60s Acidic ribosomal protein: S32-D103, M1-V30	HMMER_PFAM
					Ribosomal protein P2 signature PR00456: S90-L101	BLIMPS_PRINTS
					RIBOSOMAL PROTEIN ACIDIC 60S	BLAST_PRODOM
					PHOSPHORYLATION P2 P1 L12 MULTIGENE	
					FAMILY Y PD001928: M1-D103	
					RAT ACIDIC RIBOSOMAL PROTEIN P1	BLAST_DOMO
					DM00632	
					[P42899]1-114: M1-D103	
					[S54179]1-112: M1-D103	
					[S43109]1-113: M1-D103	
					[P41099]1-113: M1-D103	
54	7505904CD1	232	S21 S86 S94 S126 S168 S205 T112 T117 T169 T181 T228 T229	N140	Zinc finger, C2H2 type: T73-H97, Q134-H158, Y177-H201	HMMER_SMART
					Zinc finger, U1 domain: F70-Y104, D131-K165, G174-T208	HMMER_SMART
					Zinc finger, C2H2 type, domain proteins BL00028: C136-H152	BLIMPS_BLOCKS
					DSRBPZFA PD127377: P53-Q167, C75-P217	BLAST_PRODOM

Table 3

SEQ ID NO:	Incye Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
54					Zinc finger, C2H2 type, domain: C75-H97, C135-H158, C136-H158, C179-H201	MOTIFS
cont					signal_cleavage: M1-A31	
55	7509224CD1	508	S59 S181 S203 S208 S213 S235 S243 S251 S371 S377 S386 S389 S392 S403 S404 S408 S427 S437 S441 S442 S457 S502 T43 T78 T264 T395 T398	N138 N223 N369 N454 N500		SPSCAN
					RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain): V68-I137	HMMER_PFAM
					Protamine P1 proteins BL00048: R366-S392	BLIMPS_BLOCKS
					PROTEIN DNA BINDING CODED FOR BY C. ELEGANS cDNA CHROMOSOME HOMOLOG PD001830: R212-G447, K199-K417, K236-K465, P179-R375, E149-P365, D259-N493	BLAST_PRODROM
					PROTEIN TOPOISOMERASE I DNA ISOMERASE REPEAT DNA BINDING INTERMEDIATE FILAMENT HEPTAD PD000422: D259-E485, R226-D424, K183-P407, E161-P364, K303-K487, Q170-K323, E286-K506	BLAST_PRODROM
					PROTEIN REPEAT TROPOMYOSIN COILED COIL ALTERNATIVE SPLICING SIGNAL PRECURSOR CHAIN PD000023: K260-K465, K262-D479, A155-K360, R283-K506, K230-R434	BLAST_PRODROM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
55 cont					TYPE B REPEAT REPEAT DM0511  S26650 1-1203: P148-R423, R197-K461, S198-A480  P18583 113-1296: P148-R423, R165-A480	BLAST_DOMO
56	7505922CD1	274	S32 S36 S219 S240 T248 T256		Ankyrin repeat: N182-T214, L110-F142, R143-L175	HMMER_PFAM
					Ankyrin repeat: D73-F103, L110-L139, L143-L172, N182-A211	HMMER_SMART
					Ankyrin repeat proteins PF0023B: G144-E153	BLIMPS_PFAM
					PROTEIN MAJOR HISTOCOMPATIBILITY COMPLEX ENHANCER BINDING MAD3 PHOSPHORYLATION ANK REPEAT EC16/KBA PD015498: M1-Q111	BLAST_PRODROM
					PROTEIN MAJOR HISTOCOMPATIBILITY COMPLEX ENHANCER BINDING MAD3 PHOSPHORYLATION ANK REPEAT EC16/KBA PD015117: Q212-T273	BLAST_PRODROM
					ANKYRIN REPEAT DM00014  P25963 167-201: C167-L202  P25963 131-165: L131-S166  P25963 58-93: E58-I94  P25963 95-129: R95-L130	BLAST_DOMO
57	7507695CD1	527	S59 S139 S187 T54 T305 T330 T475	N67 N292	Zinc finger, C2H2 type: Y390-H412, Y502-H524, Y334-H356, Y278-H300, Y474-H496, F362-H384, F250-H272, F222-H244, F306-H328, Y418-H440, Y446-H468	HMMER_PFAM
					Zinc finger, C2H2 type, domain proteins BL00028: C336-H352	BLIMPS_BLOCKS



Table 3

SEQ ID NO.	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
57					C2H2-type zinc finger signature PR00048: P249-S262, L489-G498	BLIMPS_PRINTS
cont					Zinc-finger metal binding protein PD00066: H492-C504	BLIMPS_PRODOM
					ZINC FINGER PROTEIN ZINC FINGER METAL BINDING DNA BINDING HF.10 DIFFERENTIATION NUCLEAR PD106077: Q32-K220	BLAST_PRODOM
					PROTEIN ZINC FINGER METAL BINDING DNA BINDING ZINC FINGER PATERNALLY EXPRESSED ZN FINGER PW1 PD017719: K214-F455, G274-S515, Y334-H524, C196-I439	BLAST_PRODOM
					ZINC FINGER DNA BINDING PROTEIN METAL BINDING NUCLEAR ZINC FINGER TRANSCRIPTION REGULATION REPEAT PD000072: K276-C339, P445-C507, K248-C311	BLAST_PRODOM
					ZINC FINGER PROTEIN ZINC FINGER METAL BINDING DNA BINDING PUTATIVE REX2 TRANSCRIPTION REGULATION PD033163: K285-K416	BLAST_PRODOM
					ZINC FINGER, C2H2 TYPE, DOMAIN DM00002 IP52743 31-93: L265-H328, L377-H440 IP13682 485-515: Q493-H524 IP08042 314-358: C367-H412, C255-H300, C227-H272, C451-H496 IP13682 201-231: A209-H240	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
57 cont					Zinc finger, C2H2 type, domain: C224-H244, C252-H272, C280-H300, C308-H328, C336-H356, C364-H384, C392-H412, C420-H440, C448-H468, C476-H496	MOTIFS

Table 4

Polynucleotide SEQ ID NO./ Incyle ID/ Sequence Length	Sequence Fragments
58/7506140CB1/ 2629	1-251, 1-311, 3-240, 3-379, 12-273, 18-335, 19-286, 24-253, 29-348, 30-342, 30-822, 30-2615, 34-379, 35-379, 35-452, 41-351, 41-371, 43-372, 46-371, 55-261, 73-392, 77-371, 80-379, 84-371, 84-379, 86-374, 139-335, 141-624, 161-379, 167-396, 241-906, 255-371, 261-371, 314-361, 355-1038, 368-601, 368-851, 421-716, 465-867, 470-906, 485-904, 489-855, 551-790, 672-903, 676-1392, 677-1479, 864-1070, 892-1130, 892-1196, 892-1501, 892-1543, 892-1754, 893-1078, 893-1088, 893-1381, 902-1115, 902-1624, 907-1185, 910-1164, 917-1683, 922-1389, 925-1521, 929-1552, 930-1189, 931-1173, 931-1490, 933-1520, 934-1504, 934-1516, 934-1569, 936-1198, 936-1225, 938-1161, 938-1170, 940-1048, 941-1160, 942-1800, 943-1537, 944-1179, 948-1185, 949-1160, 949-1616, 949-1618, 952-1350, 954-1498, 955-1417, 957-1425, 957-1794, 958-1227, 958-1228, 959-1503, 967-1260, 969-1304, 1010-1613, 1012-1509, 1019-1239, 1103-1777, 1107-1542, 1126-1386, 1137-1638, 1138-1784, 1146-1700, 1180-1452, 1209-1709, 1230-1748, 1254-1906, 1289-1868, 1293-1409, 1327-1820, 1345-1598, 1366-1674, 1372-1698, 1380-1843, 1399-1899, 1405-1693, 1406-1634, 1411-1675, 1413-1717, 1419-1650, 1419-1742, 1432-2164, 1451-1697, 1474-2091, 1481-1728, 1481-1766, 1494-1781, 1501-1736, 1502-1720, 1504-2109, 1516-2048, 1538-1788, 1563-1839, 1565-1823, 1568-1842, 1569-2304, 1585-1829, 1610-2084, 1612-2233, 1616-1908, 1667-2124, 1703-1908, 1718-1908, 1746-2043, 1748-2227, 1755-2115, 1791-2081, 1804-2053, 1930-1985, 1933-1953, 1933-2006, 1940-2517, 1956-2179, 1958-2224, 1958-2233, 1958-2253, 1982-2198, 1982-2321, 1983-2201, 1983-2322, 1992-2261, 1994-2234, 2015-2270, 2021-2285, 2036-2171, 2036-2293, 2144-2432, 2155-2414, 2201-2490, 2276-2613, 2328-2614, 2343-2629, 2352-2586, 2354-2496, 2365-2629, 2366-2616, 2367-2625, 2372-2595, 2390-2501, 2390-2615, 2430-2609, 2443-2591
59/1889415CB1/ 3902	1-647, 300-984, 481-842, 481-1248, 726-1202, 989-1047, 1046-1267, 1046-1490, 1064-1490, 1122-1463, 1317-1405, 1317-1412, 1317-1490, 1317-1843, 1317-1992, 1376-1992, 1390-1992, 1425-1945, 1626-2193, 1878-2160, 1878-2162, 1878-2342, 1902-1981, 1927-2193, 2280-2549, 2291-2698, 2349-2629, 2349-2866, 2431-2955, 2462-2703, 2462-2747, 2474-3114, 2826-3536, 2911-3141, 2911-3528, 3033-3238, 3090-3695, 3113-3757, 3172-3456, 3172-3460, 3191-3580, 3249-3752, 3253-3544, 3267-3765, 3276-3520, 3310-3761, 3402-3762, 3509-3741, 3616-3902

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
60/7506047CB1/ 2993	1-482, 1-2993, 176-552, 176-659, 176-672, 176-691, 176-719, 176-730, 176-769, 176-771, 176-876, 196-905, 473-867, 473-1026, 474-752, 474-1026, 497-989, 498-956, 499-889, 536-983, 560-1016, 567-1140, 584-1159, 596-1056, 648-1159, 686-1332, 688-1159, 693-1173, 720-1165, 721-1339, 730-1270, 790-1339, 823-1436, 852-1395, 864-1339, 871-1339, 899-1340, 917-1339, 917-1340, 1023-1635, 1060-1575, 1113-1571, 1118-1635, 1199-1804, 1222-1632, 1339-1633, 1417-1633, 1631-2180, 1848-2215, 2005-2591, 2103-2805, 2369-2993, 2443-2717, 2534-2974, 2582-2967, 2771-2942
61/7505849CB1/ 679	1-252, 2-315, 2-334, 4-277, 8-411, 8-660, 9-272, 10-131, 13-131, 15-288, 15-453, 15-467, 15-517, 16-286, 18-256, 19-281, 19-306, 19-590, 23-533, 24-311, 25-597, 26-445, 26-622, 27-268, 28-408, 29-281, 33-624, 35-256, 37-232, 130-679, 131-368, 132-363, 152-621, 157-428, 158-574, 160-430, 171-659, 172-431, 175-443, 175-477, 175-483, 179-432, 181-427, 186-474, 197-375, 214-441, 214-447, 223-467, 227-313, 235-679, 243-664, 250-677, 250-679, 252-657, 258-496, 265-558, 278-659, 282-516, 294-523, 297-558, 300-561, 302-518, 308-563, 315-635, 334-653, 336-679, 342-631, 361-614, 363-491, 363-679, 376-656, 381-585, 394-610, 401-655, 404-651, 404-674, 422-599, 449-661
62/7505972CB1/ 918	1-283, 1-285, 1-294, 1-300, 1-304, 2-272, 2-283, 2-305, 3-273, 3-311, 4-219, 4-282, 4-283, 4-771, 5-223, 5-239, 5-251, 5-259, 5-283, 5-287, 6-205, 6-285, 6-300, 6-301, 6-304, 6-309, 7-178, 7-228, 7-238, 7-244, 7-252, 7-277, 7-282, 7-283, 7-284, 7-285, 7-289, 7-290, 7-293, 7-299, 7-304, 7-305, 7-306, 7-308, 7-310, 7-311, 7-511, 7-590, 7-606, 7-628, 8-254, 8-307, 8-311, 9-288, 9-300, 10-95, 10-134, 10-179, 10-219, 10-249, 10-250, 10-261, 10-269, 10-278, 10-280, 10-282, 10-283, 10-285, 10-286, 10-287, 10-288, 10-289, 10-290, 10-291, 10-293, 10-294, 10-296, 10-297, 10-298, 10-300, 10-302, 10-308, 10-309, 10-310, 10-311, 11-105, 11-222, 11-251, 11-257, 11-260, 11-261, 11-262, 11-271, 11-277, 11-292, 11-294, 11-297, 11-301, 11-310, 11-313, 11-643, 12-184, 12-206, 12-219, 12-228, 12-231, 12-238, 12-241, 12-245, 12-250, 12-253, 12-256, 12-259, 12-262, 12-266, 12-272, 12-275, 12-278, 12-280, 12-281, 12-282, 12-283, 12-284, 12-285, 12-288, 12-289, 12-294, 12-295, 12-296, 12-297, 12-298, 12-299, 12-300, 12-302, 12-307, 12-311, 13-101, 13-209, 13-264, 13-268, 13-269, 13-284, 13-298, 13-308, 13-311, 14-219, 14-254, 14-281, 14-285, 14-286, 14-297, 14-298, 14-301, 15-268, 15-290, 15-293, 15-307, 15-309, 15-310, 15-311,

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
62 cont	16-270, 16-280, 16-284, 16-291, 16-301, 16-307, 16-311, 17-223, 17-253, 17-266, 17-297, 17-311, 18-209, 19-269, 19-298, 20-193, 20-294, 20-297, 20-305, 20-311, 23-267, 23-311, 25-297, 26-243, 26-272, 26-276, 26-297, 26-305, 26-311, 28-226, 28-257, 28-277, 28-282, 28-296, 28-300, 28-302, 28-303, 28-304, 28-311, 29-311, 30-189, 32-311, 33-272, 35-311, 36-311, 40-281, 40-311, 63-115, 70-310, 94-300, 94-311, 99-260, 113-272, 113-281, 135-311, 136- 311, 152-182, 213-299, 307-480, 307-517, 307-527, 307-532, 307-543, 307-561, 307-570, 307-575, 307-579, 307- 585, 307-804, 308-544, 311-587, 314-615, 315-558, 315-572, 318-730, 319-417, 334-582, 337-604, 338-543, 340- 606, 340-620, 341-605, 342-621, 344-563, 344-575, 344-633, 347-593, 347-625, 348-605, 350-591, 351-635, 356- 599, 357-514, 357-620, 358-583, 358-590, 358-614, 358-637, 359-813, 360-604, 360-613, 363-625, 363-626, 364- 647, 364-699, 370-643, 373-651, 376-628, 381-634, 383-666, 387-662, 391-654, 391-711, 392-671, 392-672, 392- 674, 392-680, 399-679, 400-557, 400-649, 400-665, 403-664, 406-632, 406-653, 413-706, 413-804, 414-711, 416- 680, 419-670, 419-683, 423-624, 423-644, 423-645, 423-676, 431-666, 431-709, 431-726, 432-690, 432-712, 436-684, 436-700, 436-727, 441-706, 442-738, 444-918, 445-708, 445-727, 446-672, 450-708, 451-736, 453-655, 456-649, 464-666, 466-754, 468-703, 468-716, 474-734, 475-671, 478-746, 482-752, 483-739, 485-728, 488-671, 488-734, 488-750, 488-752, 488-757, 489-758, 490-699, 492-730, 492-745, 495-702, 496-754, 496-759, 500-736, 500-757, 502-742, 507-707, 508-739, 510-733, 510-744, 520-685, 520-723, 520-743, 520-753, 520-757, 549-757, 552-806, 559-735, 559-755, 563-718, 567-806, 575-733, 587-759, 629-759 1-532, 14-247, 15-294, 21-272, 21-466, 21-494, 21-585, 21-619, 21-2010, 23-302, 57-341, 58-367, 60-422, 74-323, 116-850, 202-477, 202-847, 261-501, 274-851, 282-549, 313-592, 313-878, 315-860, 331-851, 361-644, 401-680, 410-687, 438-705, 468-648, 477-725, 532-779, 542-797, 550-1060, 571-831, 589-850, 589-860, 755-967, 861-1109, 865-1109, 873-1101, 894-1603, 982-1257, 984-1480, 1002-1474, 1020-1300, 1024-1245, 1027-1592, 1038-1547, 1053-1580, 1056-1463, 1076-1478, 1079-1225, 1079-1313, 1083-1362, 1083-1648, 1086-1672, 1137-1438, 1140- 1418, 1162-1510, 1173-1449, 1177-1447, 1177-1732, 1197-1308, 1224-1619, 1228-1474, 1232-1699, 1237-1999, 1239-1492, 1239-1523, 1239-1758, 1242-1884, 1261-1876, 1267-1676, 1274-1684, 1287-1852, 1289-1536, 1289- 1544, 1290-1514, 1290-1580, 1291-1925, 1310-1700, 1319-1914, 1330-1920, 1331-1937, 1331-1992, 1340-1447, 1343-1810, 1354-1990, 1372-1823, 1375-1658, 1375-1941, 1378-1641, 1379-1929, 1383-1584, 1395-1996, 1399- 1937, 1407-1937, 1413-1937, 1424-1937, 1429-1937, 1433-1937, 1437-1864, 1437-1937, 1442-1937, 1452-1937, 1453-1937, 1454-1926, 1455-1937, 1459-1703, 1471-1937, 1473-2010, 1476-1937, 1479-1937,
63/750599 CBI/ 2010	

Table 4

Polynucleotide SEQ ID NO./ Incye ID/ Sequence Length	Sequence Fragments
63	1490-1937, 1495-2000, 1499-2008, 1521-1962, 1542-1807, 1565-1992, 1565-2010, 1599-1937, 1624-1886, 1687-1952, 1757-2001, 1757-2004, 1778-1863, 1789-2000, 1795-1870, 1805-1910, 1903-2010, 1933-2010, 1939-2004
cont	1-241, 1-475, 1-540, 20-253, 20-284, 20-4635, 25-335, 32-398, 32-460, 33-307, 44-402, 96-734, 117-542, 151-323, 160-271, 166-542, 195-542, 278-542, 876-1627, 1192-1870, 1299-1829, 1312-1870, 1382-1870, 1435-1956, 1634-1852, 1841-2059, 1994-2501, 2055-2693, 2066-2693, 2069-2693, 2072-2693, 2079-2693, 2081-2693, 2086-2488, 2086-2498, 2086-2693, 2099-2349, 2099-2374, 2099-2556, 2099-2633, 2099-2731, 2135-2693, 2198-2703, 2212-2693, 2237-2400, 2263-2732, 2276-2847, 2356-3068, 2379-2706, 2391-2681, 2400-2762, 2442-2693, 2462-3087, 2479-2672, 2523-2862, 2523-2863, 2535-3273, 2672-3032, 2694-3009, 2700-2903, 2700-2999, 2700-3261, 2700-3269, 2700-3294, 2700-3342, 2700-3353, 2726-3276, 2732-2866, 2756-3450, 2764-3357, 2779-3429, 2783-3295, 2797-3283, 2811-3455, 2825-3373, 2835-3220, 2855-3423, 2899-3455, 2919-3002, 2975-3455, 3003-3455, 3028-3355, 3052-3304, 3060-3320, 3165-3420, 3360-3901, 3360-3960, 3360-4011, 3360-4038, 3456-3696, 3456-3717, 3456-3989, 3471-4134, 3497-3785, 3510-4200, 3542-4045, 3582-4036, 3582-4038, 3594-4255, 3633-4251, 3640-4239, 3648-3927, 3649-4364, 3649-4548, 3653-3921, 3653-4126, 3677-4301, 3686-3928, 3686-3931,
64/7506003CB1/ 4677	3712-4006, 3717-4445, 3747-4497, 3771-4224, 3774-4304, 3787-4285, 3811-4368, 3827-4519, 3839-4449, 3847-4336, 3851-4142, 3852-4125, 3863-4482, 3872-4446, 3884-4501, 3905-4522, 3933-4534, 3935-4466, 3943-4531, 3948-4493, 3958-4526, 3965-4436, 3968-4169, 3969-4244, 3972-4580, 4008-4512, 4011-4650, 4021-4424, 4022-4340, 4028-4293, 4028-4629, 4070-4607, 4078-4572, 4096-4355, 4212-4612, 4218-4508, 4232-4381, 4237-4466, 4365-4580, 4365-4677, 4371-4626
65/6483977CB1/ 1110	1-252, 1-499, 1-611, 9-233, 9-458, 9-493, 9-527, 9-531, 9-544, 9-545, 9-566, 9-598, 9-617, 9-621, 9-623, 9-625, 9-639, 9-658, 9-727, 9-731, 10-225, 11-246, 11-552, 11-607, 13-219, 13-222, 13-280, 13-584, 13-584, 13-615, 14-390, 15-231, 16-270, 16-273, 16-277, 16-550, 17-436, 17-493, 18-195, 20-631, 21-236, 21-253, 21-550, 21-599, 21-605, 21-699, 22-288, 26-415, 32-286, 132-315, 234-424, 234-459, 235-491, 236-377, 236-492, 265-809, 267-501, 283-530, 285-434, 285-440, 286-539, 375-655, 382-630, 409-623, 431-661, 434-710, 490-682, 490-755, 641-826, 641-1011, 641-1026, 721-900, 721-1003, 742-1026, 840-1110, 879-1026
66/6301777CB1/ 687	1-613, 227-548, 268-687, 319-687, 448-684

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
677505976CB1/ 599	1-188, 1-237, 1-245, 1-247, 1-260, 1-261, 1-263, 1-273, 1-282, 1-307, 3-157, 3-213, 3-255, 4-253, 7-266, 13-238, 16-256, 18-242, 18-274, 20-230, 20-241, 20-256, 20-258, 20-317, 21-211, 21-278, 21-441, 26-292, 27-599, 28-283, 29-197, 33-290, 33-441, 36-181, 36-267, 38-278, 38-297, 38-343, 50-309, 52-326, 54-369, 56-166, 57-257, 57-278, 57-300, 59-171, 59-275, 60-296, 67-379, 68-299, 68-352, 68-359, 70-197, 70-302, 70-305, 70-337, 70-345, 70-384, 70-597, 70-599, 71-359, 72-346, 76-328, 76-346, 76-347, 76-439, 78-313, 78-327, 80-338, 80-346, 80-359, 81-165, 81-237, 81-256, 81-266, 81-273, 81-282, 81-322, 81-326, 81-331, 81-352, 81-360, 81-374, 82-301, 82-392, 83-357, 84-363, 84-432, 85-312, 85-379, 88-220, 88-329, 89-335, 89-344, 89-345, 89-347, 89-394, 90-213, 90-336, 90-379, 91-330, 91-401, 112-369, 113-344, 120-298, 121-369, 133-232, 159-590, 160-323, 176-589, 205-590, 232-590, 245-298, 251-332, 279-599, 300-441, 322-590, 429-590, 437-589, 437-590, 456-593, 459-586, 460-599, 482-589, 483-586, 494-590, 510-599, 519-590, 530-590
687506016CB1/ 2095	1-265, 1-411, 1-592, 1-2093, 3-510, 4-245, 9-523, 149-545, 153-565, 175-312, 176-418, 179-453, 212-556, 269-904, 285-566, 557-989, 622-872, 643-989, 655-1291, 774-1069, 787-1043, 808-1136, 886-1139, 901-1483, 911-1296, 1145-1475, 1221-1455, 1222-1471, 1226-1463, 1241-1519, 1295-1550, 1315-1835, 1323-1601, 1355-1485, 1391-1551, 1427-1631, 1475-2013, 1490-1670, 1490-1716, 1490-2095, 1497-2073, 1545-1791, 1587-1844, 1709-2095
697506086CB1/ 2843	1-241, 1-2843, 22-55, 22-87, 22-116, 22-119, 22-153, 22-157, 22-174, 22-176, 22-178, 22-181, 22-196, 22-201, 24-201, 28-201, 31-201, 49-201, 50-201, 57-201, 63-201, 64-201, 67-201, 70-201, 71-201, 79-201, 80-201, 85-201, 86-201, 93-201, 95-201, 97-201, 98-201, 99-201, 100-201, 102-201, 108-201, 113-201, 117-201, 122-201, 124-201, 125-201, 127-201, 136-201, 142-198, 142-201, 147-201, 149-201, 152-201, 166-201, 177-201, 178-201, 179-201, 196-441, 196-467, 196-712, 196-766, 196-782, 196-789, 196-807, 196-813, 196-848, 196-869, 196-948, 196-1007, 205-954, 210-577, 210-725, 210-794, 210-797, 210-840, 218-818, 236-524, 242-598, 250-908, 251-1207, 275-884, 287-559, 291-529, 305-834, 317-929, 346-817, 348-817, 348-1207, 383-927, 383-966, 383-1057, 383-1076, 383-1081, 383-1113, 383-1123, 383-1129, 383-1144, 383-1170, 383-1231, 383-1258, 384-1082, 385-1074, 385-1175, 385-1264, 396-643, 403-647, 408-1001, 413-905, 415-700, 415-722, 419-834, 419-882, 423-922, 427-1037, 442-1097, 452-1044, 452-1204, 454-924, 456-980, 468-884, 469-885, 495-1207, 512-1097, 517-720, 523-826, 523-1059, 526-826, 529-1070, 538-1180, 540-894, 542-1189, 542-1207, 547-1232, 547-1319, 549-1196,

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
69 cont	553-998, 559-1248, 596-977, 604-980, 613-1188, 615-927, 618-868, 626-1258, 632-1305, 633-861, 640-1231, 640-1271, 657-888, 659-910, 660-931, 660-1311, 660-1335, 669-915, 670-883, 672-1072, 678-1173, 684-1260, 694-1304, 694-1378, 695-1230, 698-1364, 701-834, 709-871, 711-1264, 713-1380, 716-1264, 727-972, 736-1267, 750-1328, 750-1397, 753-1258, 764-1395, 781-1471, 786-1444, 792-1347, 803-1519, 815-1378, 816-1351, 823-916, 828-1474, 870-1315, 889-1470, 930-1211, 937-1359, 992-1208, 1011-1508, 1048-1352, 1066-1310, 1069-1332, 1077-1431, 1077-1433, 1101-1280, 1109-1361, 1197-1462, 1269-1470, 1431-1711, 1493-1983, 1559-1660, 1567-1604, 1664-1970, 1664-2102, 1669-2252, 1718-1950, 1718-2106, 1718-2197, 1718-2225, 1718-2257, 1719-2403, 1727-1964, 1742-2179, 1749-2183, 1750-2298, 1755-2354, 1764-2322, 1764-2331, 1765-1920, 1765-2044, 1774-2485, 1782-2413, 1783-2015, 1783-2414, 1788-1954, 1810-2519, 1814-2106, 1815-2330, 1815-2374, 1818-2280, 1823-2081, 1823-2124, 1825-2542, 1835-2047, 1852-2430, 1863-2492, 1864-2436, 1866-2371, 1884-2453, 1886-2182, 1892-2523, 1896-2216, 1896-2585, 1898-2109, 1898-2407, 1902-2158, 1904-2085, 1908-2142, 1913-2104, 1915-2404, 1924-2600, 1930-2109, 1939-2220, 1950-2129, 1956-2563, 1959-2514, 1977-2238, 1978-2327, 1985-2242, 1985-2386, 1985-2638, 1991-2465, 1992-2446, 1995-2581, 1998-2238, 1999-2260, 2007-2274, 2007-2306, 2007-2350, 2008-2585, 2013-2597, 2014-2243, 2021-2397, 2021-2520, 2023-2218, 2025-2597, 2026-2668, 2027-2684, 2028-2517, 2030-2593, 2036-2638, 2041-2495, 2041-2603, 2041-2647, 2041-2710, 2041-2745, 2041-2777, 2041-2835, 2041-2843, 2045-2450, 2045-2545, 2045-2593, 2045-2634, 2050-2701, 2052-2313, 2053-2346, 2055-2610, 2064-2322, 2070-2648, 2070-2784, 2070-2825, 2075-2296, 2075-2325, 2077-2274, 2077-2373, 2080-2583, 2087-2843, 2094-2590, 2103-2358, 2105-2410, 2106-2609, 2107-2359, 2117-2629, 2121-2643, 2124-2368, 2125-2396, 2125-2410, 2126-2734, 2126-2795, 2133-2746, 2135-2396, 2135-2792, 2138-2675, 2140-2843, 2141-2783, 2151-2646, 2153-2758, 2161-2779, 2164-2286, 2173-2816, 2174-2396, 2174-2843, 2175-2437, 2175-2843, 2179-2472, 2182-2797, 2186-2843, 2201-2842, 2201-2843, 2205-2482, 2205-2497, 2205-2843, 2207-2764, 2215-2827, 2217-2472, 2223-2842, 2224-2843, 2225-2520, 2225-2557, 2226-2376, 2226-2404, 2226-2758, 2231-2484, 2232-2771, 2234-2811, 2240-2488, 2241-2593, 2242-2843, 2245-2843, 2248-2826, 2249-2581, 2251-2581, 2252-2561, 2254-2446, 2254-2678, 2261-2784, 2273-2843, 2278-2557, 2280-2534, 2293-2810, 2294-2579, 2315-2437, 2316-2558, 2323-2843, 2325-2843, 2335-2843, 2339-2522, 2342-2652, 2355-2670, 2367-2843, 2372-2628, 2381-2667, 2384-2843, 2385-2664, 2395-2684, 2412-2684, 2413-2573, 2417-2651, 2469-2776, 2478-2588, 2489-2742, 2490-2756, 2505-2843, 2522-2766, 2526-2730, 2528-2843, 2534-2822, 2537-2790, 2539-2827, 2650-2773, 2660-2843, 2678-2843, 2683-2843, 2709-2843, 2721-2843, 2741-2843, 2778-2843, 2786-2843



Table 4

Polynucleotide SEQ ID NO./ Incye ID/ Sequence Length	Sequence Fragments
70/4176657CB I/ 2482	1-249, 1-573, 1-665, 21-274, 26-295, 42-136, 89-727, 129-673, 161-249, 205-462, 226-918, 345-994, 390-1018, 412-1113, 426-941, 466-1146, 510-783, 638-1197, 644-960, 644-962, 653-869, 654-1306, 683-871, 690-952, 690-1046, 747-1340, 761-1027, 761-1318, 920-1449, 991-1542, 1013-1248, 1014-1260, 1015-1070, 1048-1678, 1066-1720, 1115-1788, 1155-1642, 1194-1444, 1194-1633, 1194-1642, 1204-1459, 1314-1936, 1337-1964, 1341-1939, 1358-1593, 1358-1809, 1360-2022, 1392-1868, 1496-1967, 1562-1831, 1605-1735, 1624-1770, 1720-1932, 1737-2398, 1823-2201, 1920-2179, 1942-2211, 1987-2393, 1989-2312, 1989-2393, 2056-2482, 2060-2110, 2123-2480, 2126-2482, 2384-2436, 2404-2460, 2404-2467
71/7506056CB I/ 1730	1-1730, 128-369, 129-369, 136-649, 137-423, 262-520, 370-780, 585-802, 676-786, 838-1103, 838-1309, 838-1340, 838-1438, 846-1288, 1038-1343, 1039-1347, 1049-1183, 1161-1558, 1205-1716, 1262-1730, 1289-1726, 1353-1728, 1374-1730, 1381-1730, 1382-1730, 1541-1726
72/7506185CB I/ 3119	1-288, 20-621, 20-3102, 33-560, 39-265, 50-252, 68-543, 92-377, 94-714, 124-722, 143-808, 152-500, 166-551, 166-650, 173-401, 177-683, 199-610, 204-956, 216-848, 217-857, 223-387, 225-571, 228-765, 235-607, 259-614, 259-775, 262-834, 278-797, 283-480, 321-842, 377-988, 403-1000, 422-1015, 446-841, 451-791, 452-997, 453-1135, 491-903, 499-732, 513-605, 520-859, 521-989, 536-707, 547-1054, 552-1175, 554-1076, 562-1154, 570-1187, 590-1064, 606-690, 609-1140, 611-1189, 643-755, 661-984, 675-1039, 702-1306, 713-1004, 721-1232, 746-905, 750-982, 759-963, 774-1077, 776-1120, 805-1391, 806-1240, 814-1307, 819-1050, 832-1266, 851-1131, 906-1166, 926-1306, 938-1455, 944-1158, 945-1455, 950-1769, 967-1463, 1051-1678, 1079-1315, 1141-1422, 1158-1923, 1187-1711, 1205-1525, 1222-1982, 1239-1532, 1333-1478, 1339-1616, 1380-2046, 1423-1657, 1423-1955, 1438-2179, 1449-2225, 1480-1578, 1634-2072, 1647-2280, 1658-2296, 1681-2303, 1694-2124, 1721-2224, 1731-2075, 1742-2129, 1743-2254, 1759-2060, 1759-2267, 1759-2286, 1759-2320, 1759-2351, 1759-2370, 1761-2050, 1793-2376, 1840-2430, 1860-2015, 1867-2034, 1872-2134, 1913-2286, 1954-2461, 1989-2245,
73/8096611CB I/ 1881	1997-2231, 1997-2234, 2019-2416, 2049-2322, 2066-2544, 2117-2411, 2122-2325, 2122-2341, 2139-2409, 2140-2531, 2171-2427, 2222-2371, 2228-2503, 2232-2442, 2247-2527, 2284-2504, 2368-2446, 2368-2630, 2542-3045, 2542-3099, 2544-3063, 2589-2993, 2644-2971, 2821-3116, 2822-3078, 2823-3099, 2857-3119, 2974-3086  1-670, 1-1881, 2-805, 1340-1748, 1340-1787, 1340-1805, 1340-1819, 1598-1713

Table 4

Polynucleotide SEQ ID NO./ Incye ID/ Sequence Length	Sequence Fragments
74/8174603CB1/ 1970	1-574, 1-1953, 261-577, 334-894, 742-894, 826-1439, 826-1619, 984-1647, 1001-1285, 1383-1877, 1544-1970, 1551-1720, 1704-1942, 1716-1952, 1738-1956
75/3101042CB1/ 2968	1-772, 213-804, 508-2135, 638-1408, 639-1338, 1141-1347, 1282-1339, 1308-1755, 1355-1507, 1364-1755, 1365- 1940, 1446-1680, 1446-1725, 1449-1940, 1449-2024, 1476-1924, 1510-2035, 1533-1725, 1534-1591, 1691-1743, 1691-1843, 1695-1940, 1698-1977, 1704-1924, 1722-2217, 1776-2014, 1785-1977, 1869-2024, 1972-2473, 1972- 2695, 2133-2369, 2180-2891, 2327-2962, 2342-2931, 2358-2672, 2358-2836, 2366-2962, 2422-2927, 2483-2967, 2561-2962, 2567-2786, 2659-2962, 2660-2962, 2672-2927, 2694-2823, 2698-2968, 2704-2962
76/4972035CB1/ 4497	1-584, 93-633, 286-561, 339-1060, 369-647, 417-739, 420-694, 473-752, 591-1178, 672-1052, 681-1285, 681-1300, 682-1006, 704-1170, 753-954, 791-1443, 798-1142, 854-1097, 918-1232, 963-1269, 963-1560, 963-1591, 963-1608, 967-1571, 1112-1640, 1123-1555, 1136-1678, 1298-1866, 1315-1910, 1457-2054, 1486-2258, 1541-1661, 1549- 2159, 1564-2156, 1598-1967, 1619-2128, 1625-1927, 1654-1944, 1687-2261, 1757-1980, 1775-2268, 1800-2268, 1900-2249, 1915-2350, 1929-2219, 1949-2505, 1963-2250, 2019-2557, 2108-2505, 2126-2796, 2126-2817, 2141- 2411, 2178-2466, 2188-2490, 2313-2614, 2313-2701, 2313-2742, 2313-2769, 2315-2590, 2316-2935, 2328-3000, 2334-2732, 2350-3022, 2355-2591, 2403-2689, 2437-2767, 2452-2763, 2513-2577, 2586-2856, 2652-2905, 2774- 3000, 2777-3438, 2787-3049, 2789-3047, 2789-3151, 2789-3260, 2789-3278, 2789-3306, 2790-3259, 2829-3387, 2832-3118, 2876-3480, 2914-3361, 2921-3172, 3030-3552, 3048-3438, 3074-3330, 3074-3573, 3138- 3449, 3167-3847, 3179-3810, 3213-3517, 3220-3455, 3220-3479, 3221-3476, 3254-3770, 3258-3860, 3266-3613, 3305-3857, 3309-3860, 3337-3609, 3438-3795, 3478-3982, 3489-3683, 3489-3979, 3556-3802, 3620-3953, 3665-3836, 3674-4324, 3676-3913, 3714-4003, 3740-4007, 3744-3916, 3744-4268, 3747-4180, 3749- 3961, 3749-3994, 3752-4013, 3787-4071, 3787-4086, 3811-4102, 3813-4331, 3862-4325, 3862-4330, 3889-4331, 3899-4331, 3919-4331, 3933-4331, 3972-4331, 4000-4235, 4000-4331, 4013-4497, 4016-4331, 4026-4180, 4026- 4331, 4032-4266, 4043-4320, 4048-4292

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
777506265CB1/ 513	1-480, 26-288, 27-88, 27-146, 27-153, 28-153, 31-149, 33-92, 38-136, 38-160, 38-161, 39-160, 63-469, 160-480, 164-466, 172-467, 172-471, 176-470, 180-480, 184-470, 185-460, 186-469, 187-470, 193-448, 195-418, 195-455, 195-479, 196-470, 196-485, 196-497, 200-469, 202-470, 203-448, 203-460, 203-477, 203-485, 206-468, 206-470, 209-439, 211-493, 217-470, 217-492, 225-483, 227-484, 228-483, 234-468, 239-482, 240-492, 240-494, 241-467, 242-471, 242-480, 243-483, 249-485, 251-453, 251-484, 251-495, 252-440, 254-469, 256-469, 257-487, 257-488, 260-450, 263-492, 264-470, 264-497, 266-460, 266-485, 267-513, 269-488, 270-485, 281-471, 281-475, 281-495, 284-477, 290-470, 292-513, 295-479, 295-485, 302-487, 303-495, 307-482, 308-495, 310-472, 312-450, 312-487, 317-482, 317-495, 318-480, 331-474, 334-487, 337-476, 337-485, 344-483, 366-469, 376-472, 394-484, 415-474
787506304CB1/ 509	1-504, 93-231, 93-241, 93-250, 93-344, 93-422, 94-227, 95-205, 95-250, 96-231, 96-234, 96-262, 98-185, 272-369, 283-508, 283-509, 284-509, 290-509, 299-509, 303-509, 305-501, 310-509, 314-507, 324-509, 325-454, 325-509, 331-507, 334-504, 342-509, 353-407, 356-504, 387-509, 413-509, 441-509, 451-503, 453-504
797506198CB1/ 5881	1-227, 1-5881, 19-701, 32-207, 32-405, 32-473, 32-605, 32-658, 34-434, 34-512, 36-823, 38-405, 38-435, 38-440, 38-442, 38-465, 38-507, 38-526, 39-626, 57-515, 90-716, 115-240, 134-939, 181-830, 198-805, 201-652, 229-642, 259-781, 269-369, 348-880, 375-761, 415-880, 440-880, 494-1112, 495-1073, 569-796, 583-932, 583-933, 593-1242, 626-1061, 635-973, 671-973, 737-982, 767-1471, 856-1514, 932-1529, 1044-1567, 1049-1504, 1065-1436, 1091-1571, 1125-1664, 1172-1882, 1179-1740, 1247-1500, 1275-1566, 1367-2024, 1472-2297, 1731-2294, 1769-2189, 1769-2263, 1781-2036, 1791-2011, 1791-2453, 1828-2420, 1870-2386, 1910-2165, 1910-2187, 1916-2608, 1922-2609, 1926-2609, 1929-2608, 1930-2608, 1939-2544, 1940-2605, 1943-2608, 1945-2608, 1947-2608, 1948-2608, 1953-2608, 1956-2718, 1964-2608, 1965-2605, 1968-2608, 1969-2605, 1972-2523, 1975-2608, 1976-2608, 1978-2605, 1979-2608, 1980-2608, 1985-2608, 1991-2258, 1997-2631, 1999-2608, 2001-2608, 2010-2608, 2012-2608, 2013-2605, 2014-2608, 2018-2355, 2018-2356, 2018-2364, 2018-2367, 2018-2386, 2019-2247, 2024-2362, 2031-2608, 2032-2617, 2041-2608, 2051-2608, 2051-2617, 2056-2605, 2064-2349, 2068-2608,

Table 4

Polynucleotide SEQ ID NO: / Incyte ID/ Sequence Length	Sequence Fragments
79 cont	2071-2609, 2072-2608, 2075-2348, 2078-2609, 2085-2687, 2086-2363, 2086-2608, 2090-2608, 2109-2373, 2110- 2608, 2132-2608, 2133-2408, 2163-2332, 2173-2333, 2179-2608, 2192-2617, 2193-2608, 2198-2608, 2199-2605, 2201-2608, 2233-2608, 2304-2608, 2308-2608, 2309-2576, 2314-2608, 2315-2543, 2315-3012, 2316-2876, 2318- 2608, 2318-2656, 2319-2608, 2321-2608, 2322-2608, 2324-2608, 2326-2608, 2327-2608, 2327-2609, 2331-2526, 2331-2607, 2331-2608, 2332-2608, 2333-2608, 2334-2603, 2334-2608, 2335-2608, 2336-2593, 2336-2608, 2337- 2608, 2338-2608, 2339-2608, 2341-2608, 2344-2617, 2349-2608, 2351-2618, 2352-2608, 2354-2609, 2362-2608, 2366-2608, 2367-2608, 2369-2526, 2370-2608, 2376-2608, 2377-2608, 2379-2608, 2380-2608, 2387-2656, 2391- 2609, 2394-2617, 2395-2934, 2396-2608, 2397-2608, 2400-2608, 2402-2608, 2405-3011, 2406-2608, 2407-2608, 2408-2608, 2408-3026, 2411-2608, 2415-2643, 2426-3249, 2428-2617, 2444-2608, 2456-2608, 2458-2943, 2472- 3017, 2478-2551, 2503-3073, 2535-2605, 2539-3174, 2584-3137, 2594-3030, 2608-3197, 2610-2896, 2610-2938, 2610-2939, 2610-2998, 2610-3061, 2610-3201, 2610-3219, 2611-3182, 2615-3174, 2617-2896, 2618-2809, 2618-2953, 2618-3057, 2618-3154, 2618-3198, 2619-2774, 2619-2790, 2619-2812, 2619-2814, 2619- 2822, 2619-2824, 2619-2854, 2619-2855, 2619-2857, 2619-2859, 2619-2867, 2619-2875, 2619-2878, 2619-2879, 2619-2880, 2619-2883, 2619-2884, 2619-2887, 2619-2889, 2619-2896, 2619-2898, 2619-2901, 2619-2904, 2619- 2913, 2619-2929, 2619-2932, 2619-2974, 2619-3001, 2619-3009, 2619-3052, 2619-3091, 2619-3094, 2619-3106, 2619-3125, 2619-3126, 2619-3135, 2619-3154, 2619-3163, 2619-3164, 2619-3170, 2619-3174, 2619-3176, 2619- 3183, 2619-3206, 2619-3229, 2619-3233, 2619-3234, 2619-3239, 2619-3242, 2619-3245, 2619-3256, 2619-3258, 2619-3269, 2619-3274, 2619-3278, 2619-3280, 2619-3281, 2619-3293, 2619-3294, 2619-3300, 2619-3322, 2619- 3408, 2620-3203, 2622-3256, 2624-2863, 2628-2946, 2630-3084, 2633-2963, 2637-2975, 2637-3142, 2643-2872, 2648-3027, 2651-2998, 2668-2916, 2698-3260, 2729-3405, 2754-3364, 2761-3043, 2783-3182, 2814-3490, 2835- 3089, 2847-3082, 2857-3073, 2862-3182, 2881-3514, 2915-3502, 2922-3161, 2922-3317, 2929-3171, 2942-3079, 2945-3311, 2995-3530, 3017-3294, 3030-3618, 3030-3627, 3030-3695, 3037-3338, 3037-3577,

Table 4

Polynucleotide SEQ ID NO./ Incye ID/ Sequence Length	Sequence Fragments
79 cont	<p>3055-3268, 3055-3522, 3055-3695, 3055-3699, 3063-3184, 3075-3411, 3124-3507, 3135-3403, 3158-3486, 3174-3441, 3182-3414, 3182-3681, 3188-3776, 3193-3348, 3195-3497, 3244-3769, 3245-3497, 3246-3768, 3259-3542, 3260-3947, 3269-3617, 3275-3469, 3300-3542, 3342-3613, 3353-4008, 3358-3954, 3358-4068, 3405-3677, 3407-3685, 3424-3670, 3446-3848, 3474-3886, 3476-3730, 3480-3687, 3484-3772, 3492-3812, 3494-3756, 3501-4109, 3523-3809, 3524-3765, 3537-4116, 3539-3798, 3551-3824, 3553-3681, 3580-3844, 3589-4232, 3598-3865, 3604-4133, 3641-4201, 3643-3882, 3683-3919, 3705-4157, 3713-3890, 3731-4045, 3740-4186, 3740-4365, 3744-4360, 3744-4363, 3752-4015, 3776-4120, 3777-4016, 3778-4064, 3790-4062, 3791-4062, 3821-4090, 3821-4111, 3821-4138, 3830-4112, 3837-4037, 3848-4095, 3848-4123, 3863-4162, 3863-4421, 3866-4504, 3882-4146, 3910-4150, 3920-4503, 3924-4225, 3926-4165, 3933-4200, 3937-4543, 3944-4171, 3945-4287, 3951-4146, 3970-4201, 3993-4285, 3993-4608, 4014-4541, 4043-4348, 4050-4289, 4051-4305, 4051-4354, 4055-4576, 4057-4673, 4064-4658, 4068-4688, 4083-4363, 4094-4740, 4105-4788, 4108-4708, 4118-4746, 4128-4366, 4128-4376, 4130-4659, 4138-4384, 4143-4614, 4143-4744, 4146-4411, 4148-4340, 4154-4265, 4158-4541, 4166-4803, 4169-4421, 4169-4717, 4172-4404, 4172-4410, 4172-4417, 4190-4648, 4199-4411, 4205-4492, 4212-4736, 4235-4736, 4243-4525, 4250-4513, 4253-4886, 4257-4736, 4264-4725, 4268-4521, 4272-4577, 4276-4379, 4283-4547, 4283-4736, 4285-4563, 4285-4899, 4291-4546, 4291-4822, 4298-4736, 4302-4734, 4302-4736, 4309-4736, 4312-4577, 4315-4694, 4322-4736, 4327-4597, 4327-4617, 4327-4736, 4331-4851, 4335-4594, 4337-4593, 4337-4596, 4353-4543, 4361-4658, 4370-4483, 4370-4610, 4373-4735, 4373-4745, 4383-4639, 4389-4673, 4391-4736, 4393-4643, 4394-5043, 4400-4636, 4403-4745, 4407-4890, 4408-4512, 4411-4709, 4419-4810, 4422-4810, 4423-4675, 4423-4708, 4424-4657, 4424-4737, 4431-4703, 4431-4898, 4435-4981, 4438-4728, 4441-4809, 4448-4718, 4459-4757, 4465-4678, 4465-4974, 4465-5057, 4468-4682, 4470-4615, 4472-4714, 4486-4734, 4488-4541, 4488-4735, 4498-4750, 4498-4786, 4498-5093, 4501-4809, 4524-4763, 4524-4770, 4524-4797, 4528-4810, 4529-4892, 4535-5020, 4541-4777, 4557-4796, 4572-5242, 4592-5261, 4603-4861, 4612-4813, 4625-4866, 4625-4933,</p>

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
79 cont	4632-5271, 4668-4810, 4672-4736, 4689-4967, 4693-4953, 4697-4948, 4716-5105, 4721-4993, 4735-5304, 4738- 4966, 4738-5023, 4744-4998, 4744-5000, 4744-5030, 4745-4971, 4748-5042, 4757-4924, 4763-5083, 4764-4996, 4769-5051, 4779-5054, 4781-4892, 4798-5044, 4801-5063, 4810-5079, 4815-5053, 4815-5078, 4839-5182, 4839- 5303, 4850-5536, 4851-5434, 4852-5132, 4853-5196, 4853-5334, 4853-5352, 4870-5114, 4887-5136, 4888-5295, 4905-5504, 4909-5160, 4909-5185, 4909-5461, 4916-5169, 4921-5504, 4926-5370, 4942-5222, 4943-5230, 4946- 5442, 4947-5233, 4947-5564, 4948-5255, 4950-5201, 4950-5221, 4969-5464, 4985-5805, 5000-5518, 5002-5250, 5031-5246, 5044-5163, 5044-5294, 5044-5301, 5059-5342, 5066-5352, 5068-5258, 5071-5306, 5099-5345, 5110- 5393, 5110-5723, 5111-5316, 5111-5690, 5114-5700, 5119-5337, 5134-5217, 5141-5398, 5151-5382, 5151-5414, 5151-5750, 5152-5722, 5160-5846, 5162-5398, 5164-5476, 5176-5406, 5176-5671, 5179-5470, 5203-5451, 5206- 5872, 5216-5701, 5221-5568, 5222-5381, 5226-5434, 5229-5606, 5239-5870, 5240-5863, 5246-5528, 5249-5523, 5263-5536, 5268-5498, 5269-5747, 5280-5527, 5290-5427, 5303-5566, 5309-5583, 5312-5539, 5313-5576, 5315-5487, 5316-5596, 5316-5845, 5316-5862, 5319-5556, 5324-5585, 5339-5611, 5346-5654, 5347- 5872, 5350-5583, 5370-5856, 5371-5509, 5374-5644, 5385-5618, 5385-5671, 5391-5618, 5391-5822, 5401-5806, 5402-5616, 5403-5848, 5407-5861, 5409-5616, 5413-5863, 5414-5860, 5414-5863, 5421-5706, 5421-5860, 5422- 5737, 5422-5881, 5423-5627, 5423-5637, 5423-5692, 5423-5863, 5428-5860, 5429-5881, 5437-5683, 5439-5881, 5440-5859, 5440-5863, 5442-5863, 5443-5763, 5445-5860, 5446-5752, 5448-5860, 5453-5715, 5453-5753, 5454- 5881, 5455-5881, 5456-5706, 5460-5668, 5460-5881, 5461-5860, 5468-5859, 5468-5881, 5471-5857, 5472-5767, 5474-5881, 5475-5721, 5477-5863, 5478-5733, 5478-5881, 5479-5860, 5480-5853, 5481-5715, 5481-5740, 5481- 5881, 5488-5748, 5492-5881, 5493-5857, 5494-5855, 5496-5763, 5500-5881, 5507-5856, 5517-5881, 5522-5863, 5524-5860, 5525-5705, 5528-5881, 5534-5815, 5534-5824, 5536-5825, 5542-5855, 5542-5856, 5549-5881, 5550- 5881, 5553-5860, 5555-5881, 5556-5770, 5556-5819, 5556-5860, 5557-5860, 5599-5856, 5601-5856, 5603-5863, 5607-5861, 5617-5881, 5623-5847, 5628-5860, 5686-5856, 5705-5881, 5708-5860, 5759-5860, 5769-5863, 5770-5881

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
80/1381261CBI/ 5712	<p>1-313, 1-434, 83-589, 299-1174, 345-523, 416-747, 416-848, 416-1048, 416-1096, 429-747, 452-1115, 489-1063, 517-1156, 535-1041, 575-824, 598-809, 632-981, 651-1214, 676-1222, 680-1322, 682-1398, 779-1345, 793-1372, 804-1372, 842-1405, 852-1137, 852-1404, 858-1236, 887-988, 892-1484, 898-1326, 908-1295, 972-1541, 989-1351, 1017-1478, 1024-1544, 1087-1696, 1148-1299, 1184-1739, 1220-1904, 1234-1405, 1239-1973, 1257-1979, 1271-1428, 1290-1837, 1491-2029, 1512-2094, 1668-2217, 1856-2429, 2007-2298, 2101-2691, 2102-2628, 2147-2633, 2207-2509, 2207-2536, 2230-2505, 2266-2524, 2268-2892, 2270-2484, 2463-3006, 2482-3107, 2496-3010, 2500-2836, 2516-2997, 2537-2989, 2565-2999, 2595-3085, 2639-3167, 2771-3143, 2792-3148, 2802-3179, 2830-3104, 2842-3161, 2906-3033, 2906-3117, 2906-3167, 2906-3283, 2906-3341, 2906-3507, 2906-3548, 2909-3457, 2913-3119, 2913-3344, 2921-3263, 2939-3238, 2951-3247, 2959-3370, 2994-3250, 2995-3622, 3035-3608, 3049-3296, 3056-3704, 3104-3400, 3104-3426, 3129-3432, 3131-3669, 3131-3688, 3163-3508, 3191-3430, 3226-3490, 3238-3490, 3245-3584, 3272-3533, 3274-3548, 3279-3616, 3309-3521, 3330-3613, 3331-3577,</p> <p>3351-3633, 3363-3580, 3390-3622, 3390-3634, 3443-3548, 3459-3637, 3473-3650, 3498-3724, 3498-4039, 3512-3762, 3512-3786, 3519-3932, 3519-3969, 3522-3610, 3522-3774, 3522-4051, 3551-3585, 3588-3851, 3667-3920, 3673-3946, 3712-3882, 3715-3986, 3719-4005, 3719-4162, 3732-3976, 3732-3994, 3734-3932, 3734-4021, 3744-3984, 3760-4267, 3763-4491, 3799-4470, 3818-4125, 3824-4144, 3830-4351, 3837-4128, 3841-4124, 3854-4455, 3861-4000, 3861-4122, 3866-4142, 3879-4036, 3879-4105, 3879-4163, 3895-4268, 3917-4207, 3917-4221, 3922-4172, 3935-4193, 3935-4472, 3938-4495, 3952-4287, 3952-4490, 3956-4247, 3965-4240, 3968-4241, 3971-4253, 3976-4278, 3978-4266, 3981-4266, 3992-4644, 3997-4628, 4005-4306, 4006-4280, 4023-4315, 4032-4514, 4033-4305, 4034-4264, 4034-4287, 4044-4507, 4046-4705, 4055-4302, 4059-4309, 4061-4338, 4066-4313, 4067-4351, 4079-4327, 4084-4289, 4088-4292, 4088-4639, 4095-4683, 4100-4549, 4105-4622, 4108-4358, 4115-4325, 4115-4691, 4119-4341, 4119-4640, 4122-4677, 4138-4718, 4149-4400, 4151-4424, 4159-4570, 4166-4300, 4167-4716, 4168-4481, 4180-4767, 4188-4478, 4188-4482, 4193-4438, 4197-4685, 4209-4423, 4213-4536,</p>

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
80 cont	4216-4511, 4232-4521, 4239-4458, 4256-4520, 4256-4833, 4265-4701, 4269-4885, 4274-4640, 4275-4593, 4276-4501, 4287-4549, 4287-4772, 4287-4786, 4293-4809, 4295-4712, 4297-4552, 4297-4929, 4303-4742, 4308-4531, 4321-4592, 4338-4713, 4339-4539, 4339-4596, 4339-4611, 4339-4630, 4345-4905, 4361-4882, 4363-4613, 4374-4642, 4374-4649, 4375-4670, 4388-4912, 4390-4965, 4401-4540, 4407-4946, 4418-4699, 4422-4547, 4424-4687, 4446-4929, 4452-4751, 4452-5010, 4452-5042, 4453-4714, 4458-4705, 4461-4942, 4471-4942, 4476-4942, 4477-4785, 4483-5025, 4488-4706, 4492-4952, 4495-4942, 4499-4742, 4501-4966, 4504-4952, 4506-4782, 4506-4931, 4508-4952, 4509-4778, 4512-4945, 4513-4945, 4515-4942, 4518-4952, 4521-4945, 4524-4944, 4529-4823, 4531-4918, 4532-4872, 4542-4827, 4548-4851, 4556-4851, 4559-4804, 4567-4811, 4569-4802, 4573-4843, 4574-4829, 4578-4952, 4578-5049, 4581-4942, 4586-4839, 4587-4944, 4588-4860, 4601-4926, 4601-5021, 4604-4818, 4604-5016, 4610-5048, 4613-5021, 4631-4868, 4637-4917, 4649-4990, 4659-4942, 4678-4940, 4680-4816, 4683-4942, 4690-4952, 4705-5021, 4710-4998, 4721-4964, 4755-4940, 4760-4952, 4770-5025, 4934-5198, 4958-5214, 4969-5222, 4977-5197, 4977-5207, 5001-5139, 5008-5633, 5030-5163, 5069-5659, 5101-5341, 5101-5575, 5126-5596, 5148-5442, 5159-5666, 5169-5356, 5174-5636, 5179-5712, 5196-5609, 5203-5477, 5207-5666, 5223-5679, 5238-5691, 5243-5685, 5244-5686, 5255-5684, 5258-5691, 5258-5694, 5259-5694, 5261-5684, 5265-5691, 5266-5531, 5267-5500, 5270-5691, 5272-5681, 5278-5506, 5280-5526, 5281-5508, 5281-5540, 5281-5684, 5284-5683, 5289-5681, 5290-5654, 5290-5684, 5293-5694, 5294-5694, 5295-5683, 5302-5513, 5303-5512, 5303-5671, 5303-5687, 5304-5644, 5305-5552, 5305-5670, 5307-5553, 5308-5571, 5319-5562, 5319-5653, 5324-5545, 5330-5712, 5338-5451, 5338-5577, 5340-5699, 5343-5701, 5346-5712, 5351-5601, 5357-5601, 5357-5710, 5362-5684, 5367-5691, 5368-5691, 5370-5685, 5384-5684, 5390-5684, 5393-5496, 5418-5669, 5421-5629, 5421-5647, 5428-5694, 5446-5694, 5514-5684, 5568-5686



Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
81/6803876CBI/ 1172	1-486, 356-556, 356-560, 356-561, 356-567, 356-568, 356-575, 356-579, 356-582, 356-586, 356-587, 356-589, 356-591, 356-599, 356-601, 356-603, 356-606, 356-609, 356-676, 356-741, 356-773, 356-895, 357-620, 357-643, 358-570, 358-588, 358-589, 358-600, 358-607, 358-615, 358-618, 358-650, 358-664, 358-769, 359-572, 359-584, 359-614, 359-624, 359-626, 359-642, 360-610, 360-612, 361-572, 361-602, 361-608, 361-758, 362-600, 362-753, 363-575, 363-592, 363-595, 363-597, 363-623, 363-644, 364-653, 364-778, 366-576, 366-584, 366-593, 366-613, 366-619, 366-631, 366-640, 366-687, 366-778, 367-651, 368-556, 368-604, 368-640, 369-677, 370-591, 371-601, 373-524, 376-625, 376-636, 376-663, 376-682, 377-627, 384-634, 386-635, 386-638, 386-644, 386-668, 386-685, 387-671, 389-651, 390-650, 390-674, 391-626, 391-648, 394-679, 395-640, 396-624, 396-645, 396-650, 396-656, 397-638, 397-653, 397-656, 397-681, 397-686, 398-688, 404-753, 416-615, 416-636, 416-654, 416-664, 416-678, 416-694, 416-725, 417-683, 418-668, 418-676, 418-677, 418-682, 418-685, 419-657, 420-659, 422-611, 422-688, 422-708, 422-718, 424-606, 424-676, 424-711, 425-666, 429-681, 429-720, 429-730, 430-710, 431-680, 431-685, 431-693, 433-741, 442-743, 443-696, 447-617, 452-743, 456-772, 466-624, 486-743, 507-720, 512-778, 516-703, 529-778, 534-938, 534-1148, 535-883, 567-817, 600-832, 614-841, 648-933, 665-874, 665-877, 713-972, 728-933, 728-996, 731-985, 737-1156, 740-1005, 740-1156, 742-1151, 765-1019, 782-1155, 837-869, 843-1169, 850-1156, 861-1083, 869-1115, 885-1170, 905-1154, 918-1172, 930-1155, 934-1157, 934-1170, 942-1156, 945-1172, 952-1154, 953-1172, 955-1123, 955-1154, 956-1171, 960-1155, 966-1167, 970-1172, 974-1168, 975-1172, 978-1172, 980-1172, 1004-1172, 1011-1172, 1015-1171, 1128-1171
82/7506281CBI/ 3424	1-288, 57-3418, 189-979, 189-1040, 401-993, 626-1062, 1165-1704, 1325-1754, 1333-2012, 1338-2182, 1374-1596, 1377-1620, 1400-2060, 1416-1652, 1525-1712, 1552-2285, 1563-1835, 1581-1841, 1742-1956, 1766-2587, 1796-2398, 1917-2173, 1917-2185, 1971-2249, 2054-2987, 2075-2400, 2109-2532, 2167-2987, 2169-2987, 2172-2987, 2178-2987, 2187-2987, 2190-2987, 2191-2987, 2210-2987, 2235-2987, 2236-2987, 2260-2935, 2280-2987, 2294-2830, 2308-2984, 2309-2988, 2311-2987, 2317-2986, 2325-2988, 2341-2831, 2352-2572, 2361-3154, 2369-3072, 2489-3071, 2520-3120, 2534-3166, 2541-2806, 2581-2893, 2601-2924, 2681-2918, 2698-2918, 2712-2950, 2712-2962, 2733-2867, 2739-2918, 2740-3144, 2777-3261, 2789-3029, 2796-2918, 2797-2918, 2798-2918, 2808-3424, 2815-2918, 2824-3082, 2836-3399, 2836-3418, 2868-2918, 2875-2916, 2882-3176, 2883-3293, 2958-3252, 2973-3267, 3004-3265, 3039-3176, 3046-3275, 3095-3371, 3174-3424, 3251-3366

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
83/7506175CB1/ 1882	<p>1-584, 23-283, 23-401, 27-284, 27-287, 27-1872, 33-264, 47-343, 47-543, 52-335, 52-338, 53-303, 56-306, 56-321, 57-611, 73-543, 73-594, 76-627, 81-279, 81-606, 84-340, 84-367, 88-367, 88-442, 91-353, 91-368, 91-387, 93-307, 102-395, 102-612, 120-583, 187-450, 245-829, 252-556, 288-535, 293-571, 293-578, 293-592, 293-597, 293-609, 293-619, 296-598, 316-564, 342-464, 361-937, 553-810, 582-822, 630-863, 630-1065, 630-1079, 630-1089, 630-1097, 630-1098, 631-1340, 632-878, 633-1065, 633-1095, 637-848, 637-864, 637-1037, 638-1059, 641-946, 644-1065, 647-929, 647-1097, 650-1088, 659-1065, 662-1097, 670-1089, 673-1089, 676-885, 682-1089, 684-1065, 700-1095, 711-996, 712-1089, 713-1095, 723-1097, 726-1097, 731-1089, 761-1005, 765-1095, 775-1079, 781-1025, 789-1065, 794-1065, 835-1097, 839-1048, 839-1078, 839-1089, 846-1079, 853-1065, 863-1065, 886-1089, 887-1097, 900-1675, 908-1067, 931-1358, 931-1369, 933-1207, 933-1308, 933-1318, 933-1330, 933-1358, 933-1472, 934-1324, 936-1417, 939-1641, 988-1065, 994-1065, 1013-1659, 1014-1657, 1044-1354, 1089-1340, 1089-1529, 1089-1571, 1090-1255, 1090-1269, 1090-1282, 1090-1466, 1090-1503, 1090-1508, 1090-1514,</p> <p>1090-1535, 1090-1578, 1090-1591, 1090-1592, 1091-1600, 1093-1515, 1093-1599, 1094-1354, 1096-1381, 1096-1474, 1098-1501, 1098-1515, 1098-1558, 1098-1569, 1098-1572, 1098-1592, 1098-1604, 1098-1683, 1104-1717, 1108-1430, 1108-1455, 1108-1520, 1108-1553, 1108-1560, 1108-1566, 1109-1200, 1109-1514, 1109-1573, 1112-1385, 1112-1617, 1118-1352, 1121-1446, 1121-1454, 1121-1503, 1121-1515, 1121-1540, 1121-1543, 1121-1560, 1121-1562, 1121-1574, 1121-1576, 1121-1578, 1121-1581, 1121-1582, 1128-1203, 1130-1468, 1130-1507, 1130-1543, 1130-1557, 1130-1585, 1130-1587, 1130-1588, 1134-1292, 1134-1561, 1138-1568, 1139-1375, 1143-1415, 1143-1810, 1145-1574, 1151-1641, 1152-1553, 1166-1423, 1177-1822, 1186-1335, 1198-1860, 1199-1439, 1199-1469, 1203-1502, 1203-1828, 1204-1786, 1208-1438, 1209-1458, 1212-1488, 1216-1432, 1216-1454, 1216-1490, 1221-1467, 1223-1497, 1238-1840, 1245-1469, 1248-1861, 1265-1535, 1267-1457, 1270-1666, 1271-1695, 1271-1773, 1278-1550, 1280-1866, 1289-1864, 1306-1808, 1307-1860, 1307-1862, 1350-1565, 1369-1630, 1369-1862, 1379-1882, 1381-1863, 1382-1880, 1388-1615, 1391-1851, 1396-1692, 1396-1810, 1396-1845,</p>

Table 4

Polynucleotide SEQ ID NO./ Incye ID/ Sequence Length	Sequence Fragments
83 cont	1419-1658, 1420-1700, 1422-1866, 1427-1653, 1427-1702, 1427-1882, 1429-1872, 1432-1684, 1432-1687, 1448-1864, 1451-1695, 1461-1719, 1461-1742, 1464-1868, 1482-1779, 1488-1761, 1491-1868, 1492-1798, 1492-1866, 1493-1727, 1502-1882, 1504-1866, 1508-1864, 1512-1866, 1513-1880, 1518-1863, 1518-1882, 1519-1779, 1524-1866, 1529-1868, 1531-1849, 1532-1868, 1544-1825, 1553-1806, 1555-1872, 1572-1866, 1577-1827, 1589-1862, 1600-1854, 1600-1861, 1600-1862, 1602-1862, 1603-1882, 1605-1820, 1605-1862, 1606-1882, 1615-1834, 1617-1866, 1619-1882, 1622-1882, 1629-1882, 1630-1833, 1630-1882, 1633-1882, 1635-1882, 1657-1872, 1663-1882, 1670-1862, 1671-1864, 1673-1862, 1676-1866, 1681-1868, 1683-1862, 1684-1882, 1710-1882, 1711-1881, 1730-1882, 1731-1882, 1779-1860
84/7506303CB1/ 2223	1-2118, 89-506, 89-552, 89-559, 89-822, 89-853, 109-821, 268-1068, 333-985, 374-1005, 392-1065, 426-762, 451-901, 456-1072, 462-1064, 466-1075, 564-769, 566-727, 580-1105, 591-1321, 595-1253, 620-1194, 628-1310, 643-1063, 660-1235, 680-1364, 696-1245, 711-1325, 731-1330, 753-1387, 766-1176, 784-1359, 809-1679, 819-1439, 837-1054, 837-1561, 841-1183, 868-1470, 885-1099, 889-1575, 936-1562, 946-1660, 953-1505, 955-1672, 967-1669, 985-1560, 993-1677, 998-1489, 1032-1679, 1034-1655, 1038-1605, 1082-1679, 1232-1554, 1232-1641, 1247-1382, 1251-1680, 1293-1651, 1420-2049, 1606-2223, 1638-1983, 2026-2056, 2135-2165
85/7353336CB1/ 956	1-467, 13-467, 21-467, 37-467, 42-467, 48-467, 56-467, 59-467, 67-467, 71-467, 74-467, 75-467, 82-467, 84-467, 88-467, 90-467, 91-467, 92-467, 99-823, 100-467, 101-467, 102-467, 103-467, 104-467, 104-777, 105-467, 106-467, 106-941, 107-467, 108-467, 109-467, 110-467, 112-467, 113-467, 114-467, 115-467, 116-467, 117-467, 119-467, 120-425, 120-467, 121-467, 122-467, 123-467, 124-467, 125-467, 126-467, 127-467, 128-467, 129-425, 129-467, 130-467, 130-894, 131-467, 132-467, 133-467, 134-467, 135-467, 136-467, 137-467, 139-467, 141-467, 142-467, 144-467, 147-467, 149-467, 152-467, 155-467, 156-467, 164-467, 167-467, 174-467, 192-452, 207-464, 209-467, 212-467, 213-467, 220-467, 224-467, 228-467, 232-467, 238-467, 240-467, 456-494, 456-506, 456-597, 456-607, 456-624, 456-637, 456-660, 456-665, 456-718, 456-733, 456-738, 456-740, 456-743, 456-756, 456-762, 456-763, 456-764, 456-765, 456-772, 456-775, 456-776, 456-777, 456-778, 456-782, 456-783, 456-784, 456-785, 456-786, 456-787, 456-788, 456-789, 456-790, 456-791, 456-792, 456-793, 456-794, 456-795, 456-796, 456-797, 456-798, 456-799, 456-800, 456-801, 456-802, 456-803, 456-807, 456-809, 456-812, 456-823, 456-824,

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
85 cont	456-825, 456-826, 456-827, 456-828, 456-829, 456-834, 456-836, 456-837, 456-841, 456-843, 456-844, 456-850, 456-863, 456-867, 456-869, 456-870, 456-871, 456-872, 456-873, 456-874, 456-875, 456-878, 456-879, 456-880, 456-886, 456-888, 456-890, 456-891, 456-893, 456-894, 456-896, 456-900, 456-902, 456-904, 456-905, 456-906, 456-907, 456-924, 456-927, 456-929, 456-931, 456-932, 456-935, 456-936, 456-938, 456-941, 456-942, 456-943, 456-944, 456-945, 456-954, 456-956, 498-597, 498-882
86/3001652CB1/ 2633	1-630, 1-2612, 11-321, 26-291, 45-467, 53-321, 53-588, 53-767, 87-344, 109-320, 109-470, 109-507, 109-634, 109-740, 109-746, 109-827, 122-281, 159-621, 160-359, 164-637, 164-1003, 215-776, 319-485, 333-993, 425-1024, 430-1024, 447-1029, 454-646, 457-1024, 474-1024, 489-1003, 498-1024, 540-1024, 566-1024, 576-1024, 583-772, 620-960, 626-1024, 647-1024, 686-1016, 694-1026, 704-792, 734-1281, 735-759, 735-764, 735-772, 735-774, 735-776, 735-786, 739-770, 739-774, 745-1024, 750-1024, 752-1027, 761-1024, 865-1281, 874-927, 874-928, 874-932, 874-942, 874-946, 874-997, 907-942, 907-959, 907-1040, 907-1043, 907-1081, 944-1039, 958-1011, 958-1038, 958-1081, 970-1011, 987-1024, 987-1032, 987-1037, 987-1038, 1021-1197, 1024-1170, 1025-1170, 1025-1197, 1027-1170, 1028-1170, 1032-1262, 1032-1285, 1033-1170, 1033-1197, 1033-1262, 1033-1285, 1057-1285, 1058-1197, 1087-1170, 1105-1197, 1108-1285, 1146-2385, 1199-1285, 1630-1748, 1630-1754, 1630-1790, 1630-1791, 1630-1829, 1631-1739, 1634-1829, 1685-1829, 2030-2547, 2113-2262, 2113-2288, 2116-2543, 2117-2426, 2117-2486, 2117-2494, 2117-2546, 2117-2548, 2117-2555, 2117-2556, 2117-2557, 2117-2561, 2117-2562, 2117-2569, 2117-2574, 2117-2578, 2119-2429, 2119-2478, 2119-2575, 2120-2575, 2124-2262, 2125-2288, 2125-2570, 2125-2574, 2125-2593, 2125-2633, 2135-2559, 2135-2589, 2135-2592, 2141-2526, 2143-2570, 2150-2288, 2161-2588, 2169-2540, 2169-2545, 2179-2288, 2188-2325, 2197-2288, 2210-2597, 2290-2574, 2293-2481, 2300-2574, 2316-2597, 2323-2590, 2501-2560

Table 4

Polynucleotide SEQ ID NO./ Incye ID/ Sequence Length	Sequence Fragments
87/1689128CB1/ 4521	1-453, 1-457, 1-690, 18-686, 34-690, 104-690, 134-420, 134-744, 134-752, 134-759, 134-806, 134-815, 145-877, 193-690, 288-877, 325-635, 325-817, 325-1144, 341-527, 379-877, 412-1097, 447-1141, 582-1010, 669-1114, 718-1349, 847-1560, 857-1504, 885-1431, 907-1369, 907-1502, 907-1519, 913-1548, 943-1624, 949-1602, 972-1395, 990-1285, 1149-1609, 1170-1817, 1173-1841, 1206-1930, 1221-1919, 1247-1623, 1286-1413, 1296-1478, 1338-1867, 1374-2062, 1384-1844, 1464-2049, 1482-1983, 1516-2046, 1564-2195, 1629-2216, 1643-1855, 1678-2029, 1679-1842, 1694-2310, 1709-2074, 1769-2378, 1854-2378, 1889-2626, 2210-2785, 2332-2541, 2338-2619, 2338-2898, 2573-3139, 2770-2970, 2770-3166, 2857-3342, 2879-3057, 2925-3159, 3036-3387, 3084-3354, 3101-3210, 3121-3660, 3183-3594, 3221-3335, 3224-3893, 3419-3904, 3421-3658, 3434-3730, 3434-3820, 3449-3736, 3489-3758, 3503-4236, 3564-4264, 3591-4292, 3616-4236, 3617-4210, 3631-3927, 3651-4292, 3698-4223, 3708-3986, 3708-4161, 3708-4246, 3775-4238, 3778-4217, 3786-4243, 3789-4221, 3811-4068, 3826-4264, 3849-4298, 3854-4313, 3884-4281, 3888-4281, 3890-4282, 3893-4272, 3896-4281, 3899-4279, 3914-4279, 3917-4273, 3928-4233, 3928-4280, 3935-4274, 3944-4047, 3975-4282, 3987-4281, 4036-4238, 4045-4422, 4059-4511, 4075-4431, 4075-4445, 4075-4521, 4087-4281, 4111-4281, 4141-4422
88/2362969CB1/ 3907	1-612, 13-154, 25-154, 25-161, 25-242, 26-161, 26-213, 26-220, 26-242, 28-232, 31-175, 32-242, 34-242, 37-242, 46-242, 53-242, 93-242, 143-242, 167-242, 182-233, 225-877, 240-360, 263-803, 266-981, 275-821, 311-905, 317-886, 355-830, 374-592, 374-664, 374-915, 374-919, 374-927, 378-1098, 383-1113, 393-949, 440-1046, 444-1056, 474-1005, 496-1117, 511-1020, 511-1180, 513-1064, 539-1168, 550-896, 563-1192, 599-1220, 600-1033, 604-914, 610-978, 615-891, 631-867, 639-1114, 669-1235, 669-1349, 680-887, 683-876, 725-1309, 791-1386, 791-1419, 818-1356, 836-1258, 842-1382, 859-1425, 891-1365, 909-1365, 918-1368, 925-1382, 991-1370, 1064-1405, 1160-1376, 1182-1407, 1189-1858, 1227-1393, 1234-1382, 1237-1838, 1242-1382, 1247-1382, 1391-1603, 1391-2050, 1391-2063, 1450-2092, 1683-2158, 1684-2127, 1709-2331, 1790-2156, 1790-2325, 1797-2111, 1875-2587, 1895-2085, 2064-2436, 2340-3012, 2356-2835, 2407-2987, 2605-2800, 2694-2838, 2694-3302, 2734-3001, 2757-3507, 2770-3319, 2801-3220, 2861-3122, 2861-3438, 2871-3378, 2967-3343, 3099-3509, 3123-3357, 3211-3474, 3238-3392, 3275-3445, 3422-3687, 3422-3907, 3494-3522
89/4753527CB1/ 1802	1-474, 188-597, 244-728, 431-733, 460-688, 579-733, 613-1213, 615-1802, 638-1277, 672-969, 672-1152, 694-1128, 701-1119, 710-1065, 773-1038, 810-1100, 821-1290, 831-1430, 855-1211, 982-1570, 992-1115, 992-1239, 1118-1268, 1130-1381, 1149-1569, 1158-1570, 1223-1570, 1233-1365, 1238-1570, 1259-1570, 1268-1447, 1305-1561, 1357-1570, 1362-1540, 1380-1570, 1413-1570, 1436-1570, 1448-1570, 1493-1570, 1499-1570

Table 4

Polynucleotide SEQ ID NO./ Incye ID/ Sequence Length	Sequence Fragments
90/6928688CB1/ 2469	1-589, 1-598, 1-628, 34-556, 34-604, 34-620, 34-623, 143-539, 315-562, 350-1883, 472-724, 472-754, 473-754, 496-743, 802-1097, 803-1096, 887-1180, 970-1265, 1087-1499, 1099-1702, 1159-1429, 1159-1437, 1161-1475, 1181-1760, 1289-1348, 1302-1349, 1354-1645, 1354-1679, 1354-1728, 1354-1780, 1360-1888, 1365-1654, 1414-1688, 1414-1937, 1424-1690, 1424-1851, 1433-2254, 1479-1740, 1510-1757, 1522-2092, 1528-1771, 1667-2215, 1668-2154, 1668-2238, 1668-2280, 1672-2338, 1677-2265, 1690-2294, 1698-2292, 1706-2281, 1706-2469, 1726-2208
91/7506388CB1/ 3878	1-3878, 137-392, 198-469, 208-462, 215-479, 279-779, 313-661, 316-595, 405-725, 486-761, 502-757, 502-1079, 505-698, 522-949, 551-1043, 702-1273, 747-1053, 751-1400, 794-1094, 820-968, 924-1245, 966-3878, 1046-1248, 1186-1496, 1306-1501, 1306-1521, 1332-1658, 1373-1658, 1436-1649, 1607-2134, 1805-2364, 2007-2389, 2114-2649, 2200-2383, 2269-2705, 2408-3024, 2434-3060, 2465-2971, 2619-3182, 2634-3193, 2692-3181, 2743-3216, 2850-3066, 3114-3402, 3217-3380, 3225-3683, 3230-3863, 3283-3563, 3283-3767, 3295-3868, 3306-3599, 3313-3507, 3313-3517, 3315-3750, 3319-3582, 3324-3594, 3335-3596, 3340-3835, 3354-3593, 3354-3594, 3354-3604, 3354-3624, 3366-3870, 3390-3630, 3417-3584, 3425-3675, 3426-3622, 3566-3878, 3686-3878
92/7376372CB1/ 5674	1-727, 285-780, 285-880, 515-927, 550-740, 607-5674, 757-1275, 825-1345, 863-1380
93/2754344CB1/ 3188	1-3165, 201-895, 216-878, 226-436, 239-864, 258-867, 281-898, 283-936, 296-936, 663-914, 929-1238, 1267-1527, 1267-1677, 1298-1487, 1336-1865, 1338-1543, 1348-1560, 1348-1661, 1366-1513, 1409-1758, 1486-2075, 1521-2208, 1607-1837, 1715-2048, 1715-2065, 1715-2293, 1715-2624, 1718-2023, 1774-2036, 1777-2032, 1820-2093, 1820-2094, 1854-2106, 1854-2134, 1950-2206, 2008-2261, 2065-2439, 2065-2499, 2065-2628, 2065-2633, 2065-2634, 2085-2627, 2087-2438, 2089-2467, 2089-2521, 2089-2538, 2089-2634, 2089-2688, 2089-2718, 2093-2634, 2151-2634, 2197-2714, 2202-2445, 2211-2389, 2211-2624, 2215-2498, 2235-2497, 2235-2499, 2252-2517, 2272-2456, 2273-2840, 2297-2547, 2323-2878, 2323-2896, 2323-2908, 2323-3031, 2350-2935, 2358-2610, 2363-3154, 2364-2975, 2366-2975, 2374-2648, 2380-2815, 2381-2660, 2381-2975, 2382-2975, 2391-2668, 2399-3121, 2404-2975, 2415-2956, 2420-2975, 2421-2975, 2428-2974, 2430-2668, 2439-2974, 2440-2975, 2441-2955, 2441-2972, 2441-2975, 2442-2975, 2456-3069, 2482-2975, 2484-2969, 2484-2975, 2503-2975, 2508-2975, 2515-2816, 2531-2975, 2573-2975, 2578-2793, 2585-2975, 2586-3122, 2589-2832, 2633-2907, 2635-2974, 2635-2975,

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
93 cont	2635-3066, 2635-3128, 2636-3069, 2643-3069, 2666-3188, 2686-2924, 2686-3125, 2686-3151, 2688-3183, 2716-3188, 2731-3182, 2745-3165, 2750-3165, 2771-3032, 2783-3165, 2786-3044, 2786-3059, 2786-3116, 2786-3127, 2788-3126, 2807-3099, 2861-3126, 2901-3158, 2906-3187, 2907-3144, 2916-3130, 2955-3183, 2998-3167
94/8268822CB1/ 2209	1-685, 18-740, 18-765, 18-797, 65-716, 71-679, 71-808, 75-387, 82-312, 84-299, 86-326, 89-318, 89-528, 94-615, 107-422, 112-409, 115-680, 115-733, 122-368, 122-560, 124-377, 131-432, 131-535, 131-724, 133-418, 183-958, 257-523, 258-941, 269-815, 301-803, 340-907, 344-910, 384-688, 384-853, 393-543, 422-1003, 472-1088, 492-746, 544-816, 545-746, 577-1269, 672-1389, 710-1414, 776-1415, 785-1433, 796-1035, 809-1368, 810-1077, 810-1175, 819-1085, 824-1072, 832-1306, 832-1405, 856-1110, 862-1171, 887-1525, 930-1371, 940-1313, 1036-1202, 1054-1299, 1096-1324, 1104-1319, 1116-1254, 1143-1401, 1143-1571, 1159-1389, 1245-1520, 1328-1527, 1374-1683, 1465-1685, 1506-1751, 1565-1865, 1569-1772, 1632-1864, 1649-1797, 1662-2127, 1865-2077, 1990-2196, 1990-2209
95/1814553CB1/ 2783	1-369, 75-138, 75-234, 75-316, 75-458, 75-562, 75-2783, 76-177, 77-650, 105-541, 117-639, 220-847, 350-916, 354-929, 439-1024, 464-738, 499-771, 504-931, 577-596, 600-877, 726-1293, 764-863, 821-1473, 822-1486, 826-1412, 897-1532, 949-1244, 979-1641, 1088-1624, 1094-1658, 1113-1693, 1117-1685, 1129-1893, 1187-1685, 1202-1538, 1206-1381, 1206-1744, 1209-1402, 1246-1750, 1247-1341, 1265-1410, 1324-2015, 1352-1899, 1369-1694, 1372-1745, 1416-1996, 1417-1649, 1417-1959, 1417-2020, 1418-2007, 1446-1578, 1451-2059, 1476-1693, 1486-2025, 1488-1775, 1511-1813, 1530-2173, 1542-1662, 1554-2112, 1609-2221, 1659-2270, 1676-2062, 1683-2061, 1683-2329, 1692-2164, 1698-1931, 1702-1986, 1709-2027, 1742-2173, 1767-2409, 1768-2210, 1783-2055, 1790-2233, 1790-2371, 1800-2354, 1878-2031, 1878-2135, 1878-2309, 1899-2400, 1906-2204, 1919-2204, 1919-2373, 1941-2728, 2033-2136, 2054-2621, 2088-2777, 2089-2582, 2098-2737, 2101-2378, 2101-2394, 2101-2395, 2140-2370, 2150-2577, 2174-2597, 2174-2783, 2201-2554, 2218-2712, 2234-2783, 2252-2757, 2257-2548, 2273-2740, 2278-2579, 2296-2779, 2372-2683, 2376-2712, 2383-2643, 2383-2783, 2385-2783, 2386-2753, 2398-2759, 2450-2752, 2450-2783, 2515-2742, 2518-2783, 2538-2751, 2597-2774, 2606-2779, 2656-2779, 2656-2783

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
96/71217830CB1/ 1610	1-516, 1-524, 1-538, 1-544, 1-545, 1-563, 1-612, 1-696, 50-651, 65-671, 184-793, 195-744, 477-977, 496-842, 509-1084, 525-1118, 566-766, 566-779, 566-785, 566-793, 566-794, 566-795, 566-805, 566-811, 566-819, 566-824, 566-838, 566-853, 566-864, 566-872, 566-905, 566-935, 566-981, 566-1030, 566-1076, 566-1083, 566-1105, 566-1119, 566-1126, 566-1144, 566-1172, 566-1216, 566-1239, 566-1242, 567-991, 570-845, 577-743, 578-858, 579-1040, 579-1145, 587-801, 587-803, 587-815, 587-840, 589-850, 599-866, 633-1468, 661-783, 682-913, 687-1287, 693-975, 705-1226, 716-1137, 718-949, 723-993, 738-1324, 741-1477, 756-1549, 770-1164, 790-1404, 790-1414, 804-1010, 819-1067, 819-1103, 823-1470, 844-1423, 851-1104, 877-1136, 879-1269, 888-1350, 921-1214, 922-1214, 988-1588, 991-1214, 994-1609, 996-1597, 1003-1246, 1003-1609, 1006-1295, 1018-1315, 1034-1462, 1037-1547, 1044-1596, 1047-1462, 1079-1462, 1088-1462, 1101-1610, 1126-1388, 1126-1400, 1126-1602, 1141-1501, 1147-1462, 1147-1547, 1186-1517, 1186-1547, 1190-1504, 1197-1547, 1200-1547, 1217-1517, 1223-1547, 1227-1547, 1229-1547, 1230-1547, 1231-1547, 1233-1547, 1236-1547, 1253-1513, 1255-1547, 1342-1535, 1403-1535, 1407-1535
97/7506252CB1/ 3467	1-451, 16-3450, 436-1029, 592-1147, 846-1364, 866-1147, 920-1366, 997-1261, 997-1531, 1002-1612, 1014-1512, 1096-1618, 1163-1596, 1170-1980, 1536-1799, 1552-2106, 1581-2011, 1689-1931, 1696-2280, 1699-2280, 1713-2287, 1732-1958, 1732-1994, 1732-2330, 1746-2246, 1813-2219, 1833-2045, 1967-2135, 1977-2254, 2053-2335, 2058-2356, 2145-2647, 2151-2616, 2205-2781, 2283-2505, 2296-2494, 2502-2789, 2526-2791, 2578-2805, 2683-3176, 2704-2878, 2745-3003, 2858-3060, 2858-3450, 2863-3127, 2888-3060, 2895-3210, 2920-3107, 2923-3338, 2961-3194, 3003-3248, 3004-3467, 3014-3103, 3014-3267, 3014-3323, 3019-3316, 3019-3381, 3045-3259, 3055-3328, 3072-3353, 3158-3467, 3161-3398, 3165-3301, 3217-3467, 3219-3435
98/2270608CB1/ 2392	1-2392, 189-361, 201-351, 201-376, 201-431, 201-463, 201-479, 201-748, 201-751, 201-753, 201-840, 201-865, 307-904, 308-1060, 349-1019, 420-881, 457-723, 460-998, 481-1032, 493-733, 528-1032, 533-1063, 536-806, 549-1058, 549-1204, 555-1084, 632-743, 644-826, 648-1155, 679-1170, 722-819, 734-838, 776-1367, 821-1370, 887-1379, 1019-1228, 1033-1502, 1045-1293, 1128-1664, 1196-1763, 1240-1404, 1256-1783, 1285-1540, 1285-1809, 1285-1902, 1285-1953, 1303-1543, 1303-1813, 1303-1931, 1333-1929, 1342-1588, 1344-1588, 1356-1928, 1392-1653, 1403-1545, 1424-1664, 1429-1803, 1458-1653, 1463-1953, 1469-1953, 1486-1950, 1488-1924, 1495-1745, 1496-1744, 1524-1953, 1528-1958, 1557-1956, 1585-1964, 1590-1950, 1630-1947, 1720-1950, 1742-2384, 1745-1953, 1771-1938, 1771-1953



Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
99/7502428CB1/ 1326	1-597, 52-1326, 70-697, 885-1123
100/368741CB1/ 2245	1-639, 1-778, 90-339, 90-363, 126-774, 250-490, 387-1232, 467-1282, 471-1272, 545-1290, 634-681, 634-701, 634-703, 634-759, 643-759, 648-759, 649-701, 673-1485, 677-1365, 677-1402, 677-1467, 677-1505, 677-1506, 677-1593, 684-1498, 733-831, 733-832, 797-1000, 797-1156, 797-1187, 797-1189, 797-1193, 797-1235, 797-1247, 797-1264, 797-1265, 797-1266, 797-1274, 797-1281, 797-1282, 797-1283, 797-1286, 797-1288, 797-1289, 797-1290, 797-1292, 797-1295, 797-1303, 797-1309, 797-1310, 797-1317, 797-1325, 797-1334, 797-1343, 797-1344, 797-1352, 797-1363, 797-1366, 797-1368, 797-1372, 797-1384, 797-1385, 797-1404, 797-1425, 797-1448, 798-1104, 798-1396, 798-1421, 799-1368, 807-1375, 816-1783, 817-868, 832-1366, 859-1497, 882-1517, 885-1683, 886-1011, 895-1412, 900-1721, 901-1000, 904-1517, 924-1370, 943-984, 943-995, 943-998, 943-1072, 943-1079, 943-1086, 943-1093, 943-1098, 943-1100, 943-1116, 943-1118, 943-1122, 943-1123, 943-1260, 943-1262, 945-983, 945-1536, 947-1262, 948-1262, 950-998, 951-1297, 953-1262, 964-1297, 987-1244, 998-1721, 1002-1535, 1005-1721, 1013-1739, 1028-1721, 1057-1123, 1058-1350, 1059-1262, 1070-1721, 1081-1721, 1147-1721, 1164-1249, 1164-1262, 1171-1546, 1188-2121, 1192-1375, 1195-1250, 1195-1352, 1195-1375, 1195-1502, 1195-1514, 1197-1514, 1198-1375, 1198-1514, 1200-2046, 1201-1514, 1205-1542, 1219-1464, 1222-2152, 1230-1388, 1231-1480, 1246-1501, 1309-1375, 1309-1514, 1310-1514, 1343-2165, 1416-1502, 1416-1514, 1435-1514, 1455-1514, 1470-1837, 1479-2245
101/7506379CB1/ 547	1-59, 2-462, 4-59, 5-59, 6-59, 6-290, 7-59, 23-270, 62-274, 64-262, 64-393, 68-344, 93-334, 96-530, 100-329, 106-254, 119-343, 121-234, 140-342, 142-266, 157-457, 179-299, 193-547, 209-378, 248-368, 480-506, 482-522



Table 4

Polynucleotide SEQ ID NO./ Inteyle ID/ Sequence Length	Sequence Fragments
102/7506253CB1/ 880	1-880, 24-658, 37-875, 43-148, 68-187, 209-497, 211-346, 211-395, 211-432, 211-460, 211-647, 211-669, 211-682, 211-733, 211-739, 211-767, 212-605, 213-436, 217-798, 223-579, 235-579, 254-875, 266-831, 268-579, 269-880, 282-880, 288-441, 297-880, 301-823, 306-870, 336-880, 342-614, 343-833, 348-875, 350-879, 352-616, 355-875, 358-848, 368-871, 374-880, 376-870, 383-620, 391-643, 395-880, 409-865, 413-870, 416-865, 424-684, 430-874, 430-880, 431-683, 431-880, 432-733, 432-880, 433-859, 433-876, 434-872, 439-880, 442-880, 443-666, 443-880, 445-870, 446-871, 446-874, 450-713, 450-870, 450-876, 451-871, 456-880, 461-870, 463-880, 479-880, 480-758, 487-598, 487-819, 508-879, 512-648, 514-871, 515-872, 522-871, 523-794, 523-880, 529-786, 546-739, 552-870, 557-865, 565-870, 565-880, 570-871, 603-870, 607-841, 608-877, 609-874, 621-873, 623-852, 627-880, 637-873, 656-871, 660-864, 662-860, 668-880, 670-880, 700-866, 721-880, 745-870, 750-880, 792-870
103/7506353CB1/ 813	1-222, 12-264, 12-305, 14-280, 16-349, 21-277, 23-287, 24-561, 25-268, 25-302, 25-304, 28-281, 28-800, 33-310, 35-322, 35-518, 36-314, 36-323, 44-523, 44-562, 44-567, 47-296, 59-304, 59-311, 59-315, 68-572, 73-325, 82-337, 84-228, 87-371, 89-369, 90-315, 90-341, 90-364, 91-267, 91-375, 92-333, 92-371, 96-371, 97-338, 100-321, 104-407, 104-435, 146-437, 146-572, 160-300, 160-433, 164-437, 166-649, 223-476, 236-572, 241-497, 296-534, 371-565, 378-572, 386-796, 589-766, 603-790, 606-792, 612-755, 623-804, 623-811, 626-782, 630-796, 644-808, 671-796, 680-813
104/7506372CB1/ 3014	1-164, 1-816, 1-823, 1-3014, 2-297, 2-298, 3-753, 5-429, 5-481, 5-541, 22-324, 30-302, 30-372, 32-254, 86-626, 97-654, 104-487, 104-776, 248-431, 297-547, 407-936, 481-952, 500-951, 543-954, 566-1136, 599-897, 649-1346, 650-809, 650-1310, 660-1399, 688-1585, 689-1243, 689-1387, 711-1363, 713-991, 718-972, 721-942, 759-1006, 779-1028, 795-1403, 812-1613, 818-1059, 848-1513, 881-1420, 894-1792, 933-1512, 936-1467, 943-1386, 956-1468, 958-1159, 973-1491, 974-1230, 980-1427, 980-1497, 1003-1314, 1015-1528, 1019-1604, 1020-1723, 1057-1301, 1057-1792, 1065-1462, 1074-1850, 1077-1629, 1102-1848, 1109-1655, 1125-1216, 1135-1577, 1137-1335, 1202-1339, 1224-1657, 1233-1519, 1242-1478, 1243-1426, 1260-1756, 1321-1748, 1426-1672, 1434-1588, 1450-1743, 1509-1734, 1549-1812, 1568-1843, 1583-1841, 1764-2157, 1822-2145, 1847-2046, 1847-2274, 1858-2161, 1864-2161, 1888-2152, 1888-2162, 1889-2174, 1891-2158, 1892-2016, 1926-2065, 1926-2085, 1926-2529, 1929-2158, 1931-2219, 1957-2460, 1961-2236, 1962-2233, 1964-2281, 1973-2590, 1980-2245, 1995-2242, 2019-2238, 2038-2311, 2124-2413, 2126-2522, 2154-2407, 2166-2412, 2179-2443, 2183-2654, 2191-2676, 2201-2479,

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
104 cont	2206-2927, 2246-2477, 2248-2683, 2252-2910, 2255-2509, 2300-2990, 2307-2902, 2307-2982, 2312-2855, 2319-2986, 2352-3009, 2395-2976, 2400-2983, 2401-2984, 2410-2966, 2412-2675, 2414-2670, 2436-2993, 2441-2700, 2442-2649, 2444-2985, 2478-2770, 2488-2701, 2488-3014, 2528-2998, 2530-2776, 2530-3001, 2533-2998, 2538-2991, 2539-2994, 2541-3008, 2542-3014, 2547-3003, 2550-3002, 2550-3014, 2554-3003, 2554-3008, 2557-2998, 2570-2984, 2573-3014, 2579-3014, 2581-2859, 2585-3001, 2585-3002, 2586-3002, 2587-3003, 2590-3000, 2593-3007, 2597-2991, 2599-3002, 2600-2835, 2600-2844, 2603-3014, 2605-3001, 2616-3003, 2624-2890, 2626-3002, 2628-2994, 2635-2985, 2650-3004, 2657-2988, 2675-2873, 2680-2907, 2684-2948, 2686-2969, 2705-2941, 2706-2993, 2718-2935, 2720-3000, 2721-3001, 2724-2988, 2725-2990, 2727-2986, 2727-2995, 2728-3001, 2729-3001, 2734-3012, 2735-2998, 2765-3014, 2787-3011, 2787-3014, 2789-3014, 2830-3014, 2850-3014, 2915-3009
105/7506335CB1/ 772	1-221, 1-414, 2-683, 19-163, 220-772
106/5546982CB1/ 4206	1-292, 1-669, 1-4191, 9-690, 36-767, 36-802, 37-650, 37-844, 44-747, 46-223, 50-594, 334-4206, 551-741, 566-741, 566-745, 576-1091, 917-1120, 966-1121, 966-1450, 1030-1278, 1049-1129, 1263-1561, 1424-1676, 1832-2504, 1862-2503, 1862-2515, 1950-2609, 1957-2604, 1981-2589, 2001-2575, 2001-2585, 2033-2609, 2044-2476, 2048-2526, 2048-2573, 2056-2609, 2063-2609, 2068-2448, 2069-2485, 2077-2609, 2081-2605, 2081-2609, 2088-2547, 2090-2705, 2090-2728, 2102-2550, 2144-2560, 2153-2609, 2159-2480, 2220-2737, 2222-2786, 2229-2609, 2230-2609, 2252-2607, 2270-2606, 2300-2610, 2364-2609, 2374-2940, 2382-2611, 2382-2824, 2384-2825, 2416-2687, 2416-2702, 2458-2890, 2566-2998, 3811-4204, 3837-4185, 3946-4181, 4000-4206, 4019-4206

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
107/7507432CB1/ 1971	1-602, 1-661, 2-788, 2-1962, 29-313, 29-319, 29-513, 29-1968, 30-125, 41-373, 42-482, 42-662, 52-626, 56-692, 97-309, 236-477, 238-893, 243-893, 292-519, 475-739, 634-699, 696-835, 775-886, 780-1034, 815-1086, 919-1436, 927-1180, 958-1237, 963-1181, 963-1442, 982-1494, 1002-1604, 1016-1588, 1016-1614, 1075-1346, 1081-1626, 1086-1611, 1114-1388, 1118-1722, 1123-1423, 1123-1428, 1127-1399, 1149-1418, 1201-1452, 1203-1469, 1203-1486, 1210-1820, 1212-1655, 1221-1592, 1227-1698, 1234-1737, 1236-1504, 1257-1501, 1257-1509, 1258-1660, 1267-1896, 1276-1925, 1284-1558, 1305-1860, 1324-1923, 1333-1630, 1338-1925, 1340-1605, 1345-1549, 1359-1654, 1360-1905, 1362-1917, 1422-1856, 1423-1926, 1430-1852, 1436-1666, 1437-1900, 1468-1919, 1477-1829, 1479-1971, 1489-1968, 1506-1967, 1509-1918, 1515-1971, 1540-1966, 1550-1967, 1552-1971, 1554-1966, 1560-1965, 1564-1966, 1565-1966, 1570-1798, 1572-1966, 1583-1968, 1587-1971, 1590-1971, 1592-1971, 1593-1969, 1612-1970, 1638-1970, 1642-1970, 1643-1876, 1643-1928, 1643-1970, 1646-1966, 1649-1925, 1649-1966, 1672-1865, 1698-1966
108/5639578CB1/ 1365	1-1049, 6-598, 92-1365, 489-1049, 508-1049, 775-1049
109/7509080CB1/ 5261	1-265, 1-5249, 13-272, 128-330, 532-754, 532-1048, 727-1414, 879-1432, 962-1164, 1023-1545, 1027-1545, 1214-1461, 1225-1861, 1480-1739, 1558-2147, 1650-2119, 1670-1932, 1741-1973, 2474-2686, 2868-3432, 2916-3694, 2933-3536, 2954-3474, 3155-3738, 3161-3736, 3207-3598, 3209-3506, 3231-3812, 3310-3436, 3338-3920, 3344-3995, 3362-3803, 3399-3858, 3412-4032, 3431-3647, 3435-4115, 3442-4077, 3442-4288, 3446-4115, 3451-4086, 3457-3679, 3471-3987, 3477-4026, 3489-3813, 3489-3842, 3508-4064, 3509-4110, 3510-3632, 3530-4158, 3558-4054, 3559-4062, 3563-4115, 3568-4288, 3569-4288, 3575-4111, 3582-4126, 3610-3918, 3610-4123, 3618-3837, 3632-4131, 3642-4210, 3669-3939, 3674-4357, 3687-4339, 3692-3976, 3693-3957, 3697-4002, 3698-4159, 3699-4332, 3726-4310, 3741-3974, 3758-4285, 3762-4011, 3764-4420, 3766-4246, 3778-4383, 3813-4447, 3814-4399, 3841-4100, 3844-4452, 3847-4111, 3859-4470, 3900-4463, 3914-4581, 3932-4465, 3933-4513, 3934-4197, 3937-4537, 3939-4414, 3952-4375, 3964-4507, 3990-4522, 3990-4584, 3998-4564, 3999-4699, 4027-4527, 4029-4579, 4036-4673, 4037-4549, 4045-4312, 4048-4307, 4048-4345, 4052-4287, 4054-4589, 4055-4249,

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
109 cont	4078-4372, 4082-4620, 4085-4363, 4095-4502, 4117-4724, 4126-4363, 4132-4327, 4133-4363, 4133-4367, 4140-4695, 4140-4822, 4145-4478, 4149-4785, 4151-4709, 4170-4769, 4173-4806, 4174-4763, 4197-4767, 4204-4707, 4228-4771, 4243-4811, 4252-4736, 4277-4926, 4292-4561, 4315-4941, 4321-4873, 4321-4903, 4329-4569, 4329-4977, 4332-4449, 4364-4501, 4371-4661, 4371-4914, 4376-4934, 4386-4669, 4395-4670, 4414-4919, 4414-4984, 4419-4944, 4421-4999, 4421-5024, 4471-4749, 4471-4752, 4490-4789, 4491-4770, 4513-5186, 4586-5109, 4616-5101, 4671-5195, 4733-5187, 4747-5194, 4759-5077, 4763-5194, 4801-5049, 4812-5260, 4814-5259, 4825-5259, 4835-5259, 4849-5041, 4849-5169, 4851-5134, 4853-5165, 4864-5260, 4879-5255, 4890-5097, 4890-5250, 4913-5259, 4916-5254, 4965-5253, 4981-5261, 4989-5254, 5096-5254
110/7505899CB1/ 814	1-466, 49-172, 49-195, 51-87, 51-163, 54-158, 55-292, 56-87, 57-87, 57-172, 158-456, 195-396, 195-410, 195-415, 195-442, 195-459, 195-476, 197-406, 197-454, 197-458, 197-460, 199-433, 199-459, 203-472, 203-390, 207-455, 207-460, 208-374, 208-477, 209-418, 209-459, 209-477, 210-464, 211-460, 211-473, 212-448, 212-460, 212-473, 213-405, 214-459, 215-393, 215-422, 215-431, 215-434, 215-441, 215-456, 215-458, 215-460, 216-401, 216-420, 216-422, 216-436, 216-441, 216-442, 216-459, 216-463, 217-383, 217-434, 217-440, 217-442, 217-445, 217-450, 217-454, 217-457, 217-459, 218-448, 218-473, 219-352, 219-408, 219-433, 219-454, 219-459, 219-463, 219-477, 220-426, 220-457, 222-404, 223-388, 224-408, 224-459, 224-460, 225-463, 226-425, 226-460, 227-397, 227-454, 227-459, 230-410, 230-437, 230-459, 231-477, 232-437, 232-445, 232-449, 232-459, 232-469, 233-459, 233-460, 233-463, 233-466, 233-473, 233-475, 233-477, 234-428, 234-455, 235-459, 235-460, 236-458, 236-459, 236-460, 236-467, 236-477, 239-457, 239-458, 239-473, 239-476, 240-458, 240-459, 241-453, 241-460, 241-476, 242-477, 243-472, 244-451, 244-459, 245-356, 246-451, 247-459, 249-452, 249-459, 250-379, 250-411, 250-443,

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
110 cont	<p>250-474, 251-457, 251-464, 251-466, 251-477, 252-418, 252-456, 252-463, 252-469, 252-477, 253-474, 253-477, 254-450, 254-460, 255-456, 256-460, 256-474, 258-465, 259-397, 259-460, 260-476, 260-477, 265-459, 265-460, 266-460, 268-477, 269-456, 269-461, 269-472, 269-477, 269-481, 269-497, 270-459, 270-476, 270-477, 271-459, 271-477, 272-475, 273-464, 273-476, 278-477, 279-459, 280-474, 280-475, 280-476, 282-459, 282-474, 285-468, 285-477, 286-477, 287-470, 287-477, 287-478, 290-476, 295-477, 298-457, 299-477, 300-459, 300-500, 302-477, 303-475, 303-476, 304-478, 313-437, 321-476, 323-459, 324-459, 326-477, 328-426, 329-477, 331-426, 333-474, 334-477, 336-460, 338-477, 339-477, 344-474, 345-458, 347-477, 353-463, 353-477, 355-472, 358-477, 363-452, 364-460, 366-457, 366-459, 366-460, 366-473, 367-456, 368-456, 368-457, 369-457, 369-460, 370-457, 370-477, 372-457, 372-477, 373-457, 375-460, 393-463, 396-459, 396-477, 398-460, 400-460, 460-662, 461-693, 482-693, 485-618, 485-659, 485-672, 485-682, 485-690, 485-702, 485-706, 486-566, 486-590, 486-592, 486-604, 486-609, 486-615, 486-623, 486-624, 486-628, 486-633, 486-636, 486-641, 486-658, 486-660, 486-663,</p> <p>486-667, 486-672, 486-675, 486-676, 486-677, 486-684, 486-689, 486-691, 486-692, 486-693, 486-694, 486-702, 486-709, 486-710, 486-711, 486-720, 486-721, 486-723, 486-726, 486-729, 486-731, 486-743, 486-753, 486-754, 486-814, 487-659, 487-682, 487-729, 488-629, 488-680, 488-682, 488-709, 488-712, 489-596, 489-723, 489-729, 489-744, 489-750, 489-751, 490-607, 490-693, 490-719, 490-759, 492-675, 493-710, 493-730, 494-677, 495-726, 496-711, 496-729, 496-767, 497-704, 498-689, 498-711, 498-752, 501-564, 501-587, 501-596, 501-721, 501-726, 501-750, 501-755, 502-562, 502-593, 502-596, 502-598, 502-626, 502-693, 502-696, 502-697, 502-703, 502-706, 502-708, 502-726, 502-727, 502-729, 502-736, 502-737, 502-738, 502-747, 502-751, 502-765, 503-566, 503-569, 503-596, 503-609, 503-638, 503-639, 503-662, 503-680, 503-683, 503-691, 503-692, 503-697, 503-710, 503-713, 503-715, 503-718, 503-722, 503-726, 503-727, 503-729, 503-730, 503-732, 503-735, 503-738, 503-743, 503-745, 503-746, 503-748, 503-753, 503-763, 503-809, 504-617, 504-722, 504-723, 504-726, 504-747, 504-765, 505-589, 505-590, 505-592, 505-593, 505-594, 505-596, 505-664, 505-711, 505-723, 505-742, 505-745, 506-594,</p>

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
110 cont	506-595, 506-693, 506-707, 506-710, 506-747, 506-767, 507-728, 507-755, 508-735, 508-743, 508-745, 508-765, 509-721, 510-599, 510-713, 511-716, 511-718, 512-708, 512-745, 513-730, 514-744, 514-750, 517-730, 517-745, 519-712, 520-745, 520-746, 520-767, 521-746, 521-747, 522-745, 525-649, 525-730, 525-732, 526-746, 528-745, 528-747, 529-743, 529-763, 531-747, 534-728, 536-631, 536-634, 536-742, 537-736, 540-746, 540-747, 542-746, 544-710, 544-753, 547-808, 551-712, 552-732, 552-814, 554-738, 554-743, 556-765, 557-749, 558-740, 561-746, 565-703, 565-735, 566-814, 569-747, 572-757, 574-739, 579-745, 583-712, 588-754, 606-717, 610-743, 670-767, 769-811, 778-808, 778-810, 778-814
111/7505904CB1/ 2637	1-208, 1-2637, 10-308, 14-309, 14-425, 15-324, 18-262, 23-289, 26-341, 115-548, 293-535, 293-775, 293-779, 293-809, 293-936, 293-973, 548-840, 577-1072, 577-1112, 586-1274, 589-1004, 611-1072, 692-1478, 715-971, 715-1142, 728-999, 734-828, 799-1378, 811-1133, 828-1019, 833-1153, 838-1152, 894-1487, 908-1011, 923-1174, 978-1589, 996-1228, 997-1677, 1021-1270, 1040-1333, 1075-1263, 1128-1352, 1136-1388, 1137-1396, 1185-1422, 1233-1748, 1241-1619, 1287-1577, 1287-1924, 1295-1885, 1296-1988, 1378-2051, 1416-1681, 1425-1990, 1425-1991, 1434-1991, 1445-1704, 1445-1946, 1470-1853, 1472-2102, 1504-2089, 1575-2081, 1593-2109, 1599-2097, 1639-2109, 1645-1893, 1651-2096, 1661-2116, 1665-1986, 1665-2013, 1665-2112, 1698-2096, 1710-1930, 1719-1974, 1724-1988, 1730-2112, 1753-2112, 1763-1898, 1774-2042, 1775-1921, 1779-2094, 1818-2109, 1832-2100, 1913-2156, 1948-2637, 2033-2109, 2158-2405
112/7509224CB1/ 2699	1-338, 1-340, 9-287, 10-266, 10-2353, 46-269, 46-335, 46-553, 48-501, 80-332, 80-565, 81-308, 143-595, 234-748, 388-649, 402-1004, 422-958, 448-979, 468-798, 489-720, 489-756, 558-646, 621-865, 621-881, 651-799, 672-981, 685-920, 685-960, 685-1156, 693-1070, 693-1097, 696-969, 723-1018, 739-986, 751-901, 778-1378, 787-1375, 790-893, 802-1306, 806-1238, 848-1351, 860-1508, 876-1107, 880-1385, 925-1084, 926-1340, 942-1366, 949-1526, 953-1138, 961-1222, 972-1490, 992-1480, 999-1233, 1008-1487, 1011-1172, 1020-1340, 1022-1345, 1025-1256, 1030-1285, 1036-1308, 1036-1351, 1042-1488, 1043-1364, 1074-1345, 1101-1354, 1102-1246, 1102-1313, 1105-1351, 1135-1366, 1182-1454, 1191-1495, 1283-1351, 1366-1419, 1413-1786, 1507-1753, 1601-2229, 1619-1865, 1638-1866, 1649-2455, 1652-1799, 1678-2094, 1680-2047, 1704-1978, 1708-1866, 1785-2699, 1798-2037, 1811-2202, 1811-2350, 1824-2108, 1824-2316, 1845-2426, 1863-2138, 1863-2317, 1924-2147, 1964-2333, 2018-2360, 2029-2262, 2045-2305, 2052-2340



Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
113/7505922CB1/ 1569	<p>1-243, 1-1567, 125-372, 125-416, 128-441, 130-715, 132-378, 132-816, 133-383, 134-338, 134-393, 134-497, 134-696, 134-733, 135-595, 136-861, 136-862, 137-328, 137-391, 138-862, 139-811, 140-440, 143-423, 143-574, 144-365, 144-442, 145-374, 145-526, 145-748, 147-537, 147-670, 148-397, 151-453, 151-524, 154-375, 161-352, 161-442, 169-423, 169-443, 169-552, 170-849, 173-862, 175-450, 176-538, 178-418, 192-457, 199-487, 202-435, 204-437, 205-450, 208-285, 211-508, 215-421, 216-466, 219-463, 222-748, 226-458, 229-500, 236-450, 236-509, 238-450, 243-518, 243-861, 248-511, 248-539, 254-563, 264-731, 267-636, 274-422, 276-555, 294-524, 294-575, 295-525, 303-559, 307-801, 315-843, 317-604, 321-553, 323-607, 326-582, 334-550, 339-570, 345-506, 346-862, 352-607, 356-575, 357-636, 363-689, 366-620, 366-658, 367-647, 367-656, 371-631, 371-648, 371-842, 371-847, 382-625, 383-626, 386-489, 386-663, 390-611, 394-633, 395-637, 401-587, 403-665, 420-674, 434-751, 439-671, 445-721, 449-729, 451-706, 451-727, 453-660, 453-719, 454-684, 454-715, 461-725, 463-711, 466-709, 466-724, 466-752, 468-665, 468-726, 482-765, 482-786, 483-715, 487-728, 491-698, 497-781, 498-732, 498-808,</p> <p>499-709, 501-777, 505-755, 509-789, 514-717, 514-774, 515-761, 520-761, 521-816, 529-830, 533-570, 533-799, 538-685, 538-709, 546-820, 547-774, 553-832, 558-713, 558-822, 559-851, 561-774, 561-846, 564-843, 564-859, 569-800, 574-801, 578-862, 581-857, 582-829, 586-821, 586-861, 588-862, 591-862, 600-770, 601-828, 604-791, 605-831, 609-815, 612-842, 612-850, 617-862, 628-862, 644-862, 646-859, 654-862, 663-862, 676-862, 690-862, 696-862, 710-832, 729-861, 769-846, 775-862, 775-1035, 816-1068, 860-1092, 860-1095, 860-1102, 860-1112, 860-1113, 860-1115, 860-1117, 860-1128, 860-1452, 860-1473, 860-1492, 861-1128, 864-1110, 866-1128, 866-1507, 868-1003, 870-1009, 871-1543, 873-1081, 873-1114, 874-1114, 877-1108, 877-1485, 880-1032, 881-1100, 882-1113, 883-1139, 890-1477, 900-1088, 901-1113, 902-1563, 905-1552, 905-1553, 911-1499, 913-1553, 914-1508, 922-1555, 925-1193, 925-1215, 926-1146, 927-1201, 928-1112, 935-1054, 938-1114, 939-1203, 941-1089, 941-1211, 942-1186, 949-1176, 950-1159, 950-1553, 952-1467, 953-1205, 954-1217, 956-1121, 957-1212, 958-1210, 961-1216, 963-1211, 965-1089, 965-1565, 969-1211, 972-1296, 977-1517, 978-1418, 979-1186,</p>

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
113 cont	980-1437, 982-1239, 985-1225, 985-1566, 986-1124, 987-1225, 988-1511, 989-1268, 990-1563, 993-1290, 995- 1236, 997-1169, 1001-1348, 1002-1466, 1003-1288, 1004-1099, 1004-1166, 1004-1212, 1004-1237, 1004-1306, 1006-1246, 1007-1117, 1019-1563, 1029-1244, 1031-1280, 1034-1529, 1036-1566, 1037-1354, 1040-1304, 1042- 1304, 1052-1530, 1057-1380, 1059-1268, 1061-1313, 1061-1328, 1065-1258, 1066-1298, 1066-1340, 1073-1348, 1077-1561, 1085-1310, 1085-1569, 1089-1352, 1089-1551, 1091-1565, 1092-1548, 1092-1551, 1093-1341, 1093- 1374, 1093-1547, 1095-1337, 1095-1381, 1096-1551, 1097-1308, 1103-1415, 1104-1569, 1107-1254, 1107-1378, 1107-1536, 1107-1548, 1108-1550, 1111-1554, 1116-1383, 1118-1348, 1118-1385, 1120-1551, 1122-1562, 1122- 1569, 1125-1355, 1127-1551, 1129-1548, 1131-1493, 1132-1399, 1132-1548, 1134-1569, 1135-1499, 1136-1483, 1137-1393, 1139-1548, 1140-1548, 1152-1363, 1152-1388, 1152-1449, 1154-1551, 1159-1548, 1160-1438, 1161- 1547, 1165-1553, 1170-1419, 1174-1367, 1175-1406, 1175-1454, 1180-1404, 1180-1548, 1183-1374, 1187-1422, 1189-1548, 1195-1534, 1197-1492, 1198-1442, 1199-1393, 1199-1507, 1199-1531, 1199-1533, 1200-1421, 1200-1550, 1202-1543, 1202-1562, 1203-1540, 1208-1426, 1208-1441, 1210-1445, 1213-1391, 1215- 1508, 1220-1548, 1225-1551, 1226-1551, 1229-1496, 1230-1422, 1231-1458, 1235-1509, 1235-1548, 1240-1548, 1240-1559, 1241-1463, 1244-1490, 1244-1569, 1246-1548, 1247-1490, 1247-1494, 1247-1555, 1248-1530, 1248- 1543, 1252-1503, 1252-1508, 1253-1548, 1254-1549, 1259-1541, 1260-1562, 1263-1548, 1264-1569, 1270-1548, 1271-1390, 1271-1544, 1279-1477, 1279-1500, 1289-1417, 1289-1492, 1291-1548, 1298-1433, 1304-1544, 1304- 1551, 1312-1568, 1313-1563, 1313-1564, 1315-1528, 1319-1539, 1319-1550, 1321-1497, 1321-1525, 1321-1540, 1321-1561, 1323-1560, 1324-1523, 1328-1548, 1331-1566, 1332-1534, 1332-1535, 1332-1547, 1333-1529, 1333- 1532, 1339-1541, 1339-1557, 1339-1562, 1344-1562, 1344-1564, 1345-1555, 1345-1562, 1346-1569, 1349-1569, 1359-1549, 1363-1563, 1367-1535, 1367-1552, 1371-1547, 1376-1554, 1380-1511, 1383-1557, 1389-1551, 1390- 1559, 1402-1548, 1402-1559, 1419-1475, 1419-1541, 1427-1560, 1437-1523, 1439-1569, 1441-1548, 1443-1568, 1454-1545, 1454-1569, 1455-1569, 1456-1569, 1465-1554, 1467-1569, 1470-1565, 1471-1566, 1478-1569 1-628, 1-2617, 18-578, 42-498, 42-590, 70-717, 206-1862, 337-713, 854-1063, 874-1562, 875-1056, 914-1031, 946- 1269, 950-1029, 958-1646, 1198-1521, 1201-1280, 1202-1383, 1211-1392, 1274-1483, 1538-1617, 1538-1719, 1586- 1703, 1621-1700, 1748-2046, 1748-2105, 1748-2107, 1748-2242, 1774-2319, 1919-2453, 1923-2541, 1963-2329, 1984-2528, 2000-2520, 2053-2556, 2059-2529, 2079-2528, 2083-2608, 2104-2577, 2149-2614, 2179-2619, 2180- 2617, 2199-2616, 2270-2616, 2280-2622, 2325-2852
114/7507695CB1/ 2852	

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID:	Representative Library
58	7506140CB1	MPHGNOT03
59	1889415CB1	FTUBTUE01
60	7506047CB1	KIDNFET02
61	7505849CB1	OVARTD01
62	7505972CB1	BRSTNOT23
63	7505991CB1	DRGTNOT01
64	7506003CB1	EPUNA01
65	6483977CB1	LUNGTUT03
66	6301777CB1	UTREDIT07
67	7505976CB1	ADRETUT01
68	7506016CB1	BRSTNOT04
69	7506086CB1	LIVRTUE01
70	4176657CB1	BRADDIR01
71	7506056CB1	SKINBIT01
72	7506185CB1	BRSTNOT24
73	8096611CB1	TESTNOF01
74	8174603CB1	EYERNON01
75	3101042CB1	TMLR2DT01
76	4972035CB1	THYMNOR02
77	7506265CB1	OVARNON03
78	7506304CB1	CONNTUT01
79	7506198CB1	EPUNA01
80	1381261CB1	FIBPFEN06
81	6803876CB1	PROSTUT12
82	7506281CB1	TLYMNOT03
83	7506175CB1	FIBRTXS07
84	7506303CB1	KIDETXF04
85	7353336CB1	BRAIFE01
86	3001652CB1	BMARTXE01
87	1689128CB1	BRAIFET02

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID:	Representative Library
88	2362969CB1	PROSTMY01
89	4753527CB1	TESTTUE02
90	6928688CB1	BRAFNON02
91	7506388CB1	TESTNOC01
92	7376372CB1	FTUBTUR01
93	2754344CB1	LNODNON02
94	8268822CB1	EPINOT01
95	181453CB1	THPINOT01
96	71217830CB1	BRSTTUT02
97	7506252CB1	PROSTUS19
98	2270608CB1	EPIMNON05
100	368741CB1	SYNORAT01
101	7506379CB1	MCLDXT04
102	7506253CB1	MPHGNOT03
103	7506353CB1	THP1TX04
104	7506372CB1	BRSTNOT03
105	7506335CB1	UTREDIT07
106	5546982CB1	TESTNOC01
107	7507432CB1	KIDETXJ01
108	5639578CB1	UTRSTMR01
109	7509080CB1	THYMFET03
110	7505899CB1	PANCNOT04
111	7505904CB1	PKINDNV32
112	7509224CB1	SINTNOR01
113	7505922CB1	NEUTLPT01
114	7507695CB1	DRGCNOT01

Table 6

Library	Vector	Library Description
ADRETUT01	PSPORT	Library was constructed using RNA isolated from right adrenal tumor tissue removed from a 50-year-old Turkish male during unilateral adrenalectomy. Pathology indicated a metastatic renal cell carcinoma that formed a circumscribed, spongy, hemorrhagic nodule situated in the region of the medulla. The patient presented with corticoadrenal insufficiency, incisional hernia, and non-alcoholic steato hepatitis. Patient history included renal cell carcinoma. Family history included liver cancer.
BMARTXE01	pINCY	This 5' biased random primed library was constructed using RNA isolated from treated SH-SY5Y cells derived from a metastatic bone marrow neuroblastoma, removed from a 4-year-old Caucasian female (Schering AG). The medium was MEM/HAM'S F12 with 10% fetal calf serum. After reaching about 80% confluency cells were treated with 6-Hydroxydopamine (6-OHDA) at 100 microM for 8 hours.
BRADDIR01	pINCY	Library was constructed using RNA isolated from diseased choroid plexus tissue of the lateral ventricle, removed from the brain of a 57-year-old Caucasian male, who died from a cerebrovascular accident.
BRAFNON02	pINCY	This normalized frontal cortex tissue library was constructed from 10.6 million independent clones from a frontal cortex tissue library. Starting RNA was made from superior frontal cortex tissue removed from a 35-year-old Caucasian male who died from cardiac failure. Pathology indicated moderate leptomeningeal fibrosis and multiple microinfarctions of the cerebral neocortex. Grossly, the brain regions examined and cranial nerves were unremarkable. No atherosclerosis of the major vessels was noted. Microscopically, the cerebral hemisphere revealed moderate fibrosis of the leptomeninges with focal calcifications. There was evidence of shrunken and slightly eosinophilic pyramidal neurons throughout the cerebral hemispheres. There were also multiple small microscopic areas of cavitation with surrounding gliosis scattered throughout the cerebral cortex. Patient history included dilated cardiomyopathy, congestive heart failure, cardiomegaly, and an enlarged spleen and liver. Patient medications included simethicone, Lasix, Digoxin, Colace, Zantac, captopril, and Vasotec.
		The library was normalized in two rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228 and Bonaldo et al., Genome Research (1996) 6:791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
BRAIFEF01	PCMV-ICIS	This full-length enriched library was constructed using RNA isolated from brain tissue removed from a Caucasian male fetus who was stillborn with a hypoplastic left heart at 23 weeks' gestation. Serologies were negative.
BRAIFET02	pINCY	Library was constructed using RNA isolated from brain tissue removed from a Caucasian male fetus, who was stillborn with a hypoplastic left heart at 23 weeks' gestation.

Table 6

Library	Vector	Library Description
BRSTNOT03	PSPORT1	Library was constructed using RNA isolated from diseased breast tissue removed from a 54-year-old Caucasian female during a bilateral radical mastectomy. Pathology for the associated tumor tissue indicated residual invasive grade 3 mammary ductal adenocarcinoma. Patient history included kidney infection and condyloma acuminatum. Family history included benign hypertension, hyperlipidemia and a malignant neoplasm of the colon.
BRSTNOT04	PSPORT1	Library was constructed using RNA isolated from breast tissue removed from a 62-year-old East Indian female during a unilateral extended simple mastectomy. Pathology for the associated tumor tissue indicated an invasive grade 3 ductal carcinoma. Patient history included benign hypertension, hyperlipidemia, and hematuria. Family history included cerebrovascular and cardiovascular disease, hyperlipidemia, and liver cancer.
BRSTNOT23	pINCY	Library was constructed using RNA isolated from diseased breast tissue removed from a 35-year-old Caucasian female during a bilateral reduction mammoplasty. Pathology indicated nonproliferative fibrocystic disease. Family history included type II diabetes, atherosclerotic coronary artery disease, acute myocardial infarction, hyperlipidemia, and coronary artery bypass.
BRSTNOT24	pINCY	Library was constructed using RNA isolated from diseased breast tissue removed from a 46-year-old Caucasian female during bilateral subcutaneous mastectomy. Pathology indicated nonproliferative fibrocystic disease. Family history included breast cancer and cardiovascular disease.
BRSTTUT02	PSPORT1	Library was constructed using RNA isolated from breast tumor tissue removed from a 54-year-old Caucasian female during a bilateral radical mastectomy with reconstruction. Pathology indicated residual invasive grade 3 mammary ductal adenocarcinoma. The remaining breast parenchyma exhibited proliferative fibrocystic changes without atypia. One of 10 axillary lymph nodes had metastatic tumor as a microscopic intranodal focus. Patient history included kidney infection and condyloma acuminatum. Family history included benign hypertension, hyperlipidemia, and a malignant colon neoplasm.
CONNTUT01	pINCY	Library was constructed using RNA isolated from a soft tissue tumor removed from the clival area of the skull of a 30-year-old Caucasian female. Pathology indicated chondroid chordoma with neoplastic cells reactive for keratin.

Table 6

Library	Vector	Library Description
DRGCNOT01	pINCY	Library was constructed using RNA isolated from dorsal root ganglion tissue removed from the cervical spine of a 32-year-old Caucasian male who died from acute pulmonary edema and bronchopneumonia, bilateral pleural and pericardial effusions, and malignant lymphoma (natural killer cell type). Patient history included probable cytomegalovirus infection, hepatic congestion and steatosis, splenomegaly, hemorrhagic cystitis, thyroid hemorrhage, and Bell's palsy. Surgeries included colonoscopy, large intestine biopsy, adenotomylectomy, and nasopharyngeal endoscopy and biopsy; treatment included radiation therapy.
EPIMNON05	pINCY	This normalized mammary epithelial cell tissue library was constructed from 3.28 million independent clones from an epithelial cell tissue library. Starting RNA was made from untreated mammary epithelial cell tissue removed from a 21-year-old female. The library was normalized in two rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228 and Bonaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48 -hours/round) reannealing hybridization was used.
EPIPNOT01	pINCY	Library was constructed using RNA isolated from prostatic epithelial cells removed from a 17-year-old Hispanic male.
EPTPUNA01	PSPORT1	Library was constructed using RNA isolated from untreated prostatic epithelial cell tissue removed from a 17-year-old Hispanic male. Serologies were negative.
EYERNON01	PSPORT1	This normalized pooled retina tissue library was constructed from independent clones from a pooled retina tissue library. Starting RNA was made from pooled retina tissue removed from 34 male and female donors, aged 9 to 80-years-old. The library was normalized in one round using conditions adapted from Soares et al., PNAS (1994) 91:9228-9232 and Bonaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
FIBPFEN06	pINCY	The normalized prostate stromal fibroblast tissue libraries were constructed from 1.56 million independent clones from a prostate fibroblast library. Starting RNA was made from fibroblasts of prostate stroma removed from a male fetus, who died after 26 weeks' gestation. The libraries were normalized in two rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228 and Bonaldo et al., Genome Research (1996) 6:791, except that a significantly longer (48-hours/round) reannealing hybridization was used. The library was then linearized and recircularized to select for insert containing clones as follows: plasmid DNA was prepped from approximately 1 million clones from the normalized prostate stromal fibroblast tissue libraries following soft agar transformation.

Table 6

Library	Vector	Library Description
FIBRTXS07	pINCY	This subtracted library was constructed using 1.3 million clones from a dermal fibroblast library and was subjected to two rounds of subtraction hybridization with 2.8 million clones from an untreated dermal fibroblast tissue library. The starting library for subtraction was constructed using RNA isolated from treated dermal fibroblast tissue removed from the breast of a 31-year-old Caucasian female. The cells were treated with 9CIS retinoic acid. The hybridization probe for subtraction was derived from a similarly constructed library from RNA isolated from untreated dermal fibroblast tissue from the same donor. Subtractive hybridization conditions were based on the methodologies of Swaroop et al., NAR (1991) 19:1954 and Bonaldo, et al., Genome Research (1996) 6:791.
FTUBTUE01	pINCY	This 5' biased random primed library was constructed using RNA isolated from right fallopian tube tumor tissue removed from an 85-year-old Caucasian female during bilateral salpingo-oophorectomy and hysterectomy. Pathology indicated poorly differentiated mixed endometrioid (80%) and serous (20%) adenocarcinoma of the right fallopian tube, which was confined to the mucosa without mural involvement. Endometrioid carcinoma in situ was also present. Pathology for the associated uterus tumor indicated focal endometrioid adenocarcinoma in situ and moderately differentiated invasive adenocarcinoma arising in an endometrial polyp. A metastatic endometrioid and serous adenocarcinoma was present in the cul-de-sac tumor. The patient presented with a pelvic mass and ascites. Patient history included medullary carcinoma of the thyroid and myocardial infarction. Patient medications included Nitro-Dur, Lescol, Lasix and Cardizem.
FTUBTUR01	PCDNA2.1	This random primed library was constructed using RNA isolated from fallopian tube tumor tissue removed from an 85-year-old Caucasian female during bilateral salpingo-oophorectomy and hysterectomy. Pathology indicated poorly differentiated mixed endometrioid (80%) and serous (20%) adenocarcinoma, which was confined to the mucosa without mural involvement. Endometrioid carcinoma in situ was also present. Pathology for the associated uterus tumor indicated focal endometrioid adenocarcinoma in situ and moderately differentiated invasive adenocarcinoma arising in an endometrial polyp. Metastatic endometrioid and serous adenocarcinoma was present at the cul-de-sac tumor. Patient history included medullary carcinoma of the thyroid and myocardial infarction.
KIDETXF04	PCMV-ICIS	Library was constructed using RNA isolated from a treated, transformed embryonal cell line (293-EBNA) derived from kidney epithelial tissue. The cells were treated with 5-aza-2'-deoxycytidine (5AZA) for 72 hours and Trichostatin A for 24 hours and transformed with adenovirus 5 DNA.



Table 6

Library	Vector	Library Description
KIDETXJ01	pIGEN	This random primed 5' cap isolated library was constructed using RNA isolated from a treated, transformed embryonal cell line (293-EBNA) derived from kidney epithelial tissue (Invitrogen). The cells were treated with 5-aza-2'-deoxycytidine (SAZA) for 72 hours and Trichostatin A for 24 hours and transformed with adenovirus 5 DNA.
KIDNFET02	pINCY	Library was constructed using RNA isolated from kidney tissue removed from a Caucasian male fetus, who was stillborn with a hypoplastic left heart and died at 23 weeks' gestation.
LIVRTUE01	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from liver tumor tissue removed from a 72-year-old Caucasian male during partial hepatectomy. Pathology indicated metastatic grade 2 (of 4) neuroendocrine carcinoma forming a mass. The patient presented with metastatic liver cancer. Patient history included benign hypertension, type I diabetes, prostatic hyperplasia, prostate cancer, alcohol abuse in remission, and tobacco abuse in remission. Previous surgeries included destruction of a pancreatic lesion, closed prostatic biopsy, transurethral prostatectomy, removal of bilateral testes and total splenectomy. Patient medications included Eulexin, Hytrin, Proscar, Ecotrin, and insulin. Family history included atherosclerotic coronary artery disease and acute myocardial infarction in the mother; atherosclerotic coronary artery disease and type II diabetes in the father.
LNODNON02	pINCY	This normalized lymph node tissue library was constructed from .56 million independent clones from a lymph node tissue library. Starting RNA was made from lymph node tissue removed from a 16-month-old Caucasian male who died from head trauma. Serologies were negative. Patient history included bronchitis. Patient medications included Dopamine, Dobutamine, Vancomycin, Vasopressin, Proventil, and Atarax. The library was normalized in two rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228-9932 and Bonaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
LUNGTUT03	PSPORT1	Library was constructed using RNA isolated from lung tumor tissue removed from the left lower lobe of a 69-year-old Caucasian male during segmental lung resection. Pathology indicated residual grade 3 invasive squamous cell carcinoma. Patient history included acute myocardial infarction, prostatic hyperplasia, malignant skin neoplasm, and tobacco use.

Table 6

Library	Vector	Library Description
MCLDXT04	pINCY	Library was constructed using RNA isolated from treated, derived dendritic cells from umbilical cord blood CD34+ precursor cells removed from a male. The cells were derived with granulocyte/macrophage colony stimulating factor (GM-CSF), tumor necrosis factor alpha (TNF alpha), and stem cell factor (SCF), then treated with phorbol myristate acetate (PMA), and Ionomycin. The GM-CSF was added at time 0 at 100 ng/ml, the TNF alpha was added at time 0 at 2.5 ng/ml, and the SCF was added at time 0 at 25 ng/ml. The PMA and Ionomycin were added at 13 days for five hours. Incubation time was 13 days.
MPHGNOT03	PBLUESCRIPT	Library was constructed using RNA isolated from plastic adherent mononuclear cells isolated from buffy coat units obtained from unrelated male and female donors.
NEUTLPT01	PBLUESCRIPT	Library was constructed using RNA isolated from peripheral blood granulocytes collected by density gradient centrifugation through Ficoll-Hypaque. The cells were isolated from buffy coat units obtained from unrelated male and female donors. Cells were cultured in 100 ng/ml E. coli LPS for 30 minutes, lysed in GuSCN, and spun through CsCl to obtain RNA for library construction.
OVARNON03	pINCY	This normalized ovarian tissue library was constructed from 5 million independent clones from an ovary library. Starting RNA was made from ovarian tissue removed from a 36-year-old Caucasian female during total abdominal hysterectomy, bilateral salpingo-oophorectomy, soft tissue excision, and an incidental appendectomy. Pathology for the associated tumor tissue indicated one intramural and one subserosal leiomyomata of the myometrium. The endometrium was proliferative phase. Patient history included deficiency anemia, calculus of the kidney, and a kidney anomaly. Family history included hyperlipidemia, acute myocardial infarction, atherosclerotic coronary artery disease, type II diabetes, and chronic liver disease. The library was normalized in two rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228 and Bonaldo et al., Genome Research (1996) 6:791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
OVARTDT01	pINCY	Library was constructed using RNA isolated from right ovary tissue removed from a 47-year-old Caucasian female during a total abdominal hysterectomy with bilateral salpingo-oophorectomy. Pathology for the associated tumor tissue indicated two intramural leiomyomas. The endometrium was in the secretory phase. The patient presented with abnormal blood chemistry. Patient history included infertility, abnormal blood chemistry, abnormal heart sounds, and extrinsic asthma. Family history included benign hypertension, atherosclerotic coronary artery disease, an aortic valve disorder, acute myocardial infarction, cerebrovascular disease, and pancreatic cancer.
PANCNOT04	PSPORT1	Library was constructed using RNA isolated from the pancreatic tissue of a 5-year-old Caucasian male who died in a motor vehicle accident.

Table 6

Library	Vector	Library Description
PKINDNV32	PCR2-TOPOTA	Library was constructed using pooled cDNA from different donors. cDNA was generated using mRNA isolated from pooled skeletal muscle tissue removed from ten 21 to 57-year-old Caucasian male and female donors who died from sudden death; from pooled thymus tissue removed from nine 18 to 32-year-old Caucasian male and female donors who died from sudden death; from pooled liver tissue removed from 32 Caucasian male and female fetuses who died at 18-24 weeks gestation due to spontaneous abortion; from kidney tissue removed from 59 Caucasian male and female fetuses who died at 20-33 weeks gestation due to spontaneous abortion; and from brain tissue removed from a Caucasian male fetus who died at 23 weeks gestation due to fetal demise.
PROSTMV01	pINCY	This large size-fractionated cDNA and normalized library was constructed using RNA isolated from diseased prostate tissue removed from a 55-year-old Caucasian male during closed prostatic biopsy, radical prostatectomy, and regional lymph node excision. Pathology indicated adenofibromatous hyperplasia. Pathology for the matched tumor tissue indicated adenocarcinoma Gleason grade 4 forming a predominant mass involving the left side peripherally with extension into the right posterior superior region. The tumor invaded the capsule and perforated the capsule to involve periprostatic tissue in the left posterior superior region. The left inferior posterior and left superior posterior surgical margins are positive. One left pelvic lymph node is metastatically involved. Patient history included calculus of the kidney. Family history included lung cancer and breast cancer. The size-selected library was normalized in 1 round using conditions adapted from Soares et al., PNAS (1994) 91:9228-9232 and Bonaldo et al., Genome Research (1996) 6:791.
PROSTUS19	pINCY	This subtracted prostate tumor tissue library was constructed using 2.36 million clones from a prostate tumor library and was subjected to two rounds of subtraction hybridization with 2.36 million clones from a prostate epithelium library. The starting library for subtraction was constructed using RNA isolated from prostate tumor tissue removed from a 59-year-old Caucasian male during a radical prostatectomy with regional lymph node excision. Pathology indicated adenocarcinoma (Gleason grade 3+3) involving the prostate peripherally with invasion of the capsule. Adenofibromatous hyperplasia was present. The patient presented with elevated prostate-specific antigen (PSA). Patient history included diverticulitis of colon, asbestosis, and thrombophlebitis. Family history included benign hypertension, multiple myeloma, hyperlipidemia, and rheumatoid arthritis. Subtractive hybridization conditions were based on the methodologies of Swaroop et al., NAR (1991) 19:1954 and Bonaldo, et al. Genome Research (1996) 6:791.

Table 6

Library	Vector	Library Description
PROSTUT12	pINCY	Library was constructed using RNA isolated from prostate tumor tissue removed from a 65-year-old Caucasian male during a radical prostatectomy. Pathology indicated an adenocarcinoma (Gleason grade 2+2). Adenofibromatous hyperplasia was also present. The patient presented with elevated prostate specific antigen (PSA).
SINTNOR01	PCDNA2.1	This random primed library was constructed using RNA isolated from small intestine tissue removed from a 31-year-old Caucasian female during Roux-en-Y gastric bypass. Patient history included clinical obesity.
SKINBIT01	pINCY	Library was constructed using RNA isolated from diseased skin tissue of the left lower leg. Patient history included erythema nodosum of the left lower leg.
SYNORAT01	PSPORT1	Library was constructed using RNA isolated from synovial membrane tissue removed from the elbow of a 51-year-old Asian female with rheumatoid arthritis.
TESTNOC01	PBLUESCRIPT	This large size fractionated library was constructed using RNA isolated from testicular tissue removed from a pool of eleven, 10 to 61-year-old Caucasian males.
TESTNOF01	PSPORT1	This 5' cap isolated full-length library was constructed using RNA isolated from testis tissue removed from a 26-year-old Caucasian male who died from head trauma due to a motor vehicle accident. Serologies were negative. Patient history included a hernia at birth, tobacco use (1 1/2 ppd), marijuana use, and daily alcohol use (beer and hard liquor).
TESTTUE02	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from testicular tumor removed from a 31-year-old Caucasian male during unilateral orchiectomy. Pathology indicated embryonal carcinoma forming a largely necrotic mass involving the entire testicle. Rare foci of residual testicle showed intralobular germ cell neoplasia and tumor was identified at the spermatic cord margin. The patient presented with backache. Patient history included tobacco use. Previous surgeries included a needle biopsy of testis. Patient medications included Colace and antacids.
THP1NOT01	pINCY	Library was constructed using RNA isolated from untreated THP-1 cells. THP-1 is a human promonocyte line derived from the peripheral blood of a 1-year-old Caucasian male with acute monocytic leukemia (ref: Int. J. Cancer (1980) 26:171).
THP1TXT04	pINCY	Library was constructed using RNA isolated from stimulated THP-1 cells. THP-1 is a human promonocyte line derived from the peripheral blood of a 1-year-old male (Abbott Sample) with acute monocytic leukemia (Int. J. Cancer 26 (1980):171).
THYMFET03	pINCY	Library was constructed using RNA isolated from thymus tissue removed from a Caucasian male fetus.

Table 6

Library	Vector	Library Description
THYMNOR02	pINCY	The library was constructed using RNA isolated from thymus tissue removed from a 2-year-old Caucasian female during a thymectomy and patch closure of left atrioventricular fistula. Pathology indicated there was no gross abnormality of the thymus. The patient presented with congenital heart abnormalities. Patient history included double inlet left ventricle and a rudimentary right ventricle, pulmonary hypertension, cyanosis, subaortic stenosis, seizures, and a fracture of the skull base. Family history included reflux neuropathy.
TLYMNOR03	pINCY	Library was constructed using RNA isolated from nonactivated Th1 cells. These cells were differentiated from umbilical cord CD4 T cells with IL-12 and B7-transfected COS cells.
TMLR2DT01	PBLUESCRIPT	Library was constructed using RNA isolated from non-adherent peripheral blood mononuclear cells. The blood was obtained from unrelated male and female donors. Cells from each donor were purified on Ficoll Hypaque, then co-cultured for 24 hours in medium containing normal human serum at a cell density of 2million cells/ml.
UTREDIT07	pINCY	Library was constructed using RNA isolated from diseased endometrial tissue removed from a female during endometrial biopsy. Pathology indicated in phase endometrium with missing beta 3, Type II defects.
UTRSTMR01	pINCY	Library was constructed using RNA isolated from uterine myometrial tissue removed from a 41-year-old Caucasian female during a vaginal hysterectomy. The endometrium was secretory and contained fragments of endometrial polyps. Pathology for associated tumor tissue indicated uterine leiomyoma. Patient history included ventral hernia and a benign ovarian neoplasm.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	ESTs: Probability value = 1.0E-8 or less; Full Length sequences: Probability value = 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E value = 1.0E-6; Assembled ESTs: fasta Identity = 95% or greater and Match length = 200 bases or greater; fastx E value = 1.0E-8 or less; Full Length sequences: fastx score = 100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value = 1.0E-3 or less

Table 7

Program	Description	Reference	Parameter Threshold
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM, INCY, SMART and TIGRFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	PFAM, INCY, SMART or TIGRFAM hits: Probability value = 1.0E-3 or less; Signal peptide hits: Score = 0 or greater
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score $\geq$ GCG specified "HIGH" value for that particular Prosite motif. Generally, score = 1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score = 120 or greater; Match length = 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score = 3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	

Table 7

Program	Description	Reference	Parameter Threshold
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. On Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence (AAAI) Press, Menlo Park, CA, and MIT Press, Cambridge, MA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	



Table 8

SEQ ID NO.	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
76	4972035	2427525H1	SNP00015185	8	4007	T	T	G	noncoding	n/a	n/a	n/a	n/a
77	7506265	007882H1	SNP00020849	62	182	T	T	C	I41	n/a	n/a	n/a	n/a
78	7506304	1294222H1	SNP00039835	13	108	G	G	A	noncoding	n/a	n/a	n/a	n/a
79	7506198	1002021H1	SNP00012943	12	5187	T	C	T	noncoding	0.81	0.86	0.84	0.80
81	6803876	1210089H1	SNP00066548	92	224	A	A	G	noncoding	n/a	n/a	n/a	n/a
82	7506281	2686636H1	SNP00107537	162	2873	T	T	C	noncoding	n/a	n/a	n/a	n/a
83	7506175	1856272H1	SNP00014990	186	287	T	C	T	R61	n/a	n/a	n/a	n/a
84	7506303	3937378H1	SNP00009517	105	1454	C	C	G	noncoding	n/a	n/a	n/a	n/a
85	7353336	031320H1	SNP00032569	256	562	G	G	C	noncoding	n/d	n/a	n/a	n/a
86	3001632	6712119H1	SNP00049926	153	906	T	C	T	N277	n/a	n/a	n/a	n/a
87	1689128	1910984H1	SNP00009056	131	3838	T	C	T	noncoding	n/a	n/a	n/a	n/a
88	2362969	6910906J1	SNP00047603	14	1103	A	A	C	N216	n/a	n/a	n/a	n/a
89	475327	3597648H1	SNP00112423	183	854	G	G	A	W264	n/d	n/a	n/a	n/a
91	7506388	4674120H1	SNP00024799	254	451	A	G	A	P65	0.33	n/a	n/a	n/a
92	7376372	2600624H1	SNP00060994	210	1855	G	G	A	A429	0.70	n/a	n/a	n/a
93	2754344	2745172H1	SNP00000294	270	1987	T	T	C	L562	0.27	0.08	0.10	0.23
94	8268822	3383569H1	SNP00031965	195	316	C	C	T	T48	n/d	0.99	n/d	n/d
95	1814553	1870195H1	SNP00105341	212	1687	C	C	T	H508	n/d	n/d	n/d	n/d
97	7506252	2006992H1	SNP00051840	31	2950	T	C	T	L979	n/a	n/a	n/a	n/a
100	368741	7667604H1	SNP00129354	389	386	G	A	G	S83	n/a	n/a	n/a	n/a
101	7506379	1432222H1	SNP00058616	11	210	C	C	A	P62	n/d	n/a	n/a	n/a
102	7506253	1322352H1	SNP00029137	99	276	C	C	A	D67	n/d	n/a	n/a	n/a
103	7506353	1831292H1	SNP00036352	151	661	G	G	T	noncoding	n/d	n/a	n/a	n/a
104	7506372	3166112H1	SNP00029300	157	643	C	C	T	Q172	n/a	n/a	n/a	n/a
106	5546982	5822041H1	SNP00074077	35	2097	G	G	T	S597	n/a	n/a	n/a	n/a
107	7507432	3390140H1	SNP00044438	170	1525	T	T	C	F500	n/d	n/a	n/a	n/a
109	7509080	6500745H1	SNP00015247	360	4778	T	T	C	noncoding	n/a	n/a	n/a	n/a
110	7505899	6031045H1	SNP00047347	125	181	G	C	G	K24	n/d	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
111	7505904	5836174H1	SNP00015294	241	968	C	C	T	noncoding	n/a	n/a	n/a	n/a
112	7509224	2702219H1	SNP00037452	138	1509	A	G	A	E322	n/a	n/a	n/a	n/a
113	7505922	4184173H1	SNP00013088	126	301	C	C	T	L25	0.75	0.98	0.76	0.77
114	7507695	1274378H1	SNP000003360	117	2079	G	A	G	noncoding	0.74	n/a	n/a	n/a

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:

- a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-57,
- b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-2, SEQ ID NO:13, SEQ ID NO:16, SEQ ID NO:18-20, SEQ ID NO:22-23, SEQ ID NO:31-32, SEQ ID NO:36-39, SEQ ID NO:41, SEQ ID NO:43, SEQ ID NO:46, SEQ ID NO:51, and SEQ ID NO:55,
- c) a polypeptide comprising a naturally occurring amino acid sequence at least 91% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:21, and SEQ ID NO:27,
- d) a polypeptide comprising a naturally occurring amino acid sequence at least 94% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:17, SEQ ID NO:20, and SEQ ID NO:42,
- e) a polypeptide comprising a naturally occurring amino acid sequence at least 95% identical to the amino acid sequence of SEQ ID NO:50,
- f) a polypeptide comprising a naturally occurring amino acid sequence at least 96% identical to the amino acid sequence of SEQ ID NO:19,
- g) a polypeptide comprising a naturally occurring amino acid sequence at least 98% identical to the amino acid sequence of SEQ ID NO:24,
- h) a polypeptide comprising a naturally occurring amino acid sequence at least 99% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:9, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:52, and SEQ ID NO:57,
- i) a polypeptide consisting essentially of a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:3-8, SEQ ID NO:10-12, SEQ ID NO:14-15, SEQ ID NO:24-27, SEQ ID NO:29, SEQ ID NO:34, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44-45, SEQ ID NO:47-49, SEQ ID NO:53-54, and SEQ ID NO:56,
- j) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-57, and

- k) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-57.

2. An isolated polypeptide of claim 1 comprising an amino acid sequence selected from the  
5 group consisting of SEQ ID NO:1-57.

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

10

5. An isolated polynucleotide of claim 4 comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:58-114.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a  
15 polynucleotide of claim 3.

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

20

9. A method of producing a polypeptide of claim 1, the method comprising:

- a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and  
25  
b) recovering the polypeptide so expressed.

25

10. A method of claim 9, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-57.

30

11. An isolated antibody which specifically binds to a polypeptide of claim 1.

12. An isolated polynucleotide selected from the group consisting of:

- a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:58-114,
- b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:58-67, SEQ ID NO:69-86, SEQ ID NO:88, SEQ ID NO:93, SEQ ID NO:95-96, SEQ ID NO:98, SEQ ID NO:100, SEQ ID NO:108, and SEQ ID NO:110-112,
- c) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 91% identical to the polynucleotide sequence of SEQ ID NO:94,
- d) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 93% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:106, and SEQ ID NO:114,
- e) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 94% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:87, and SEQ ID NO:103,
- f) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 96% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:89-90, SEQ ID NO:92, and SEQ ID NO:101,
- g) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 97% identical to the polynucleotide sequence of SEQ ID NO:99,
- h) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 98% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:68, SEQ ID NO:91, and SEQ ID NO:109,
- i) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 99% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:102, and SEQ ID NO:107,
- j) a polynucleotide consisting essentially of a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:97, SEQ ID NO:104-105, and SEQ ID NO:113,
- k) a polynucleotide complementary to a polynucleotide of a),
- l) a polynucleotide complementary to a polynucleotide of b),
- m) a polynucleotide complementary to a polynucleotide of c),

- n) a polynucleotide complementary to a polynucleotide of d),
  - o) a polynucleotide complementary to a polynucleotide of e),
  - p) a polynucleotide complementary to a polynucleotide of f),
  - q) a polynucleotide complementary to a polynucleotide of g),
  - 5 r) a polynucleotide complementary to a polynucleotide of h),
  - s) a polynucleotide complementary to a polynucleotide of i),
  - t) a polynucleotide complementary to a polynucleotide of j), and
  - u) an RNA equivalent of a)-t).
- 10 13. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 12.
14. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:
- 15 a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and
  - 20 b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.
15. A method of claim 14, wherein the probe comprises at least 60 contiguous nucleotides.
- 25 16. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:
- a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and
  - b) detecting the presence or absence of said amplified target polynucleotide or fragment
  - 30 thereof, and, optionally, if present, the amount thereof.
17. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable

excipient.

18. A composition of claim 17, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-57.

5

19. A method for treating a disease or condition associated with decreased expression of functional NAAP, comprising administering to a patient in need of such treatment the composition of claim 17.

10 20. A method of screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- b) detecting agonist activity in the sample.

15 21. A composition comprising an agonist compound identified by a method of claim 20 and a pharmaceutically acceptable excipient.

22. A method for treating a disease or condition associated with decreased expression of functional NAAP, comprising administering to a patient in need of such treatment a composition of  
20 claim 21.

23. A method of screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- 25 b) detecting antagonist activity in the sample.

24. A composition comprising an antagonist compound identified by a method of claim 23 and a pharmaceutically acceptable excipient.

30 25. A method for treating a disease or condition associated with overexpression of functional NAAP, comprising administering to a patient in need of such treatment a composition of claim 24.

26. A method of screening for a compound that specifically binds to the polypeptide of claim 1, the method comprising:

- a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and
- 5 b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

27. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, the method comprising:

- 10 a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,
- b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and
- c) comparing the activity of the polypeptide of claim 1 in the presence of the test  
15 compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

20 28. A method of screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

- a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,
- 25 b) detecting altered expression of the target polynucleotide, and
- c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

29. A method of assessing toxicity of a test compound, the method comprising:

- 30 a) treating a biological sample containing nucleic acids with the test compound,
- b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 12 under conditions



whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 12 or fragment thereof,

- c) quantifying the amount of hybridization complex, and
- d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

30. A method for a diagnostic test for a condition or disease associated with the expression of NAAP in a biological sample, the method comprising:

- a) combining the biological sample with an antibody of claim 11, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex, and
- b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

31. The antibody of claim 11, wherein the antibody is:

- a) a chimeric antibody,
- b) a single chain antibody,
- c) a Fab fragment,
- d) a F(ab')<sub>2</sub> fragment, or
- e) a humanized antibody.

32. A composition comprising an antibody of claim 11 and an acceptable excipient.

33. A method of diagnosing a condition or disease associated with the expression of NAAP in a subject, comprising administering to said subject an effective amount of the composition of claim 32.

34. A composition of claim 32, further comprising a label.

35. A method of diagnosing a condition or disease associated with the expression of NAAP

in a subject, comprising administering to said subject an effective amount of the composition of claim 34.

36. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 11, the method comprising:

- a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-57, or an immunogenic fragment thereof, under conditions to elicit an antibody response,
- b) isolating antibodies from the animal, and
- c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-57.

37. A polyclonal antibody produced by a method of claim 36.

15

38. A composition comprising the polyclonal antibody of claim 37 and a suitable carrier.

39. A method of making a monoclonal antibody with the specificity of the antibody of claim 11, the method comprising:

- a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-57, or an immunogenic fragment thereof, under conditions to elicit an antibody response,
- b) isolating antibody producing cells from the animal,
- c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells,
- d) culturing the hybridoma cells, and
- e) isolating from the culture monoclonal antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-57.

25  
30

40. A monoclonal antibody produced by a method of claim 39.

41. A composition comprising the monoclonal antibody of claim 40 and a suitable carrier.

42. The antibody of claim 11, wherein the antibody is produced by screening a Fab expression library.

43. The antibody of claim 11, wherein the antibody is produced by screening a recombinant immunoglobulin library.

44. A method of detecting a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-57 in a sample, the method comprising:

- a) incubating the antibody of claim 11 with the sample under conditions to allow specific binding of the antibody and the polypeptide, and
- b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-57 in the sample.

45. A method of purifying a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-57 from a sample, the method comprising:

- a) incubating the antibody of claim 11 with the sample under conditions to allow specific binding of the antibody and the polypeptide, and
- b) separating the antibody from the sample and obtaining the purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-57.

46. A microarray wherein at least one element of the microarray is a polynucleotide of claim 13.

47. A method of generating an expression profile of a sample which contains polynucleotides, the method comprising:

- a) labeling the polynucleotides of the sample,
- b) contacting the elements of the microarray of claim 46 with the labeled polynucleotides of the sample under conditions suitable for the formation of a hybridization complex, and
- c) quantifying the expression of the polynucleotides in the sample.

48. An array comprising different nucleotide molecules affixed in distinct physical locations on a solid substrate, wherein at least one of said nucleotide molecules comprises a first oligonucleotide or polynucleotide sequence specifically hybridizable with at least 30 contiguous nucleotides of a target polynucleotide, and wherein said target polynucleotide is a polynucleotide of claim 12.

5

49. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 30 contiguous nucleotides of said target polynucleotide.

50. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is  
10 completely complementary to at least 60 contiguous nucleotides of said target polynucleotide.

51. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to said target polynucleotide.

15 52. An array of claim 48, which is a microarray.

53. An array of claim 48, further comprising said target polynucleotide hybridized to a nucleotide molecule comprising said first oligonucleotide or polynucleotide sequence.

20 54. An array of claim 48, wherein a linker joins at least one of said nucleotide molecules to said solid substrate.

55. An array of claim 48, wherein each distinct physical location on the substrate contains multiple nucleotide molecules, and the multiple nucleotide molecules at any single distinct physical  
25 location have the same sequence, and each distinct physical location on the substrate contains nucleotide molecules having a sequence which differs from the sequence of nucleotide molecules at another distinct physical location on the substrate.

56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

30

57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

5 61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

10

64. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

65. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.

15 66. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.

67. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.

68. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:13.

20

69. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:14.

70. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:15.

25 71. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:16.

72. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:17.

73. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:18.

30

74. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:19.

75. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:20.

76. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:21.

77. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:22.

5 78. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:23.

79. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:24.

80. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:25.

10

81. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:26.

82. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:27.

15

83. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:28.

84. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:29.

85. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:30.

20

86. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:31.

87. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:32.

25

88. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:33.

89. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:34.

90. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:35.

30

91. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:36.

92. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:37.

93. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:38.

94. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:39.

5 95. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:40.

96. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:41.

97. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:42.

10

98. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:43.

99. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:44.

15 100. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:45.

101. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:46.

102. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:47.

20

103. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:48.

104. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:49.

25 105. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:50.

106. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:51.

107. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:52.

30

108. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:53.

109. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:54.

110. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:55.
111. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:56.
- 5 112. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:57.
113. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:58.
- 10 114. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:59.
115. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:60.
- 15 116. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:61.
117. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:62.
- 20 118. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:63.
- 25 119. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:64.
120. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:65.
- 30 121. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:66.



122. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID  
NO:67.

123. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID  
5 NO:68.

124. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID  
NO:69.

10 125. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID  
NO:70.

126. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID  
NO:71.

15 127. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID  
NO:72.

128. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID  
20 NO:73.

129. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID  
NO:74.

25 130. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID  
NO:75.

131. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID  
NO:76.

30 132. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID  
NO:77.

133. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID  
NO:78.

134. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID  
5 NO:79.

135. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID  
NO:80.

10 136. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID  
NO:81.

137. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID  
NO:82.

15 138. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID  
NO:83.

139. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID  
20 NO:84.

140. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID  
NO:85.

25 141. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID  
NO:86.

142. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID  
NO:87.

30 143. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID  
NO:88.

144. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:89.

145. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:90.

146. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:91.

147. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:92.

148. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:93.

149. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:94.

150. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:95.

151. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:96.

152. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:97.

153. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:98.

154. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:99.

155. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID  
NO:100.

156. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID  
5 NO:101.

157. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID  
NO:102.

10 158. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID  
NO:103.

159. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID  
NO:104.  
15

160. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID  
NO:105.

161. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID  
20 NO:106.

162. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID  
NO:107.

25 163. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID  
NO:108.

164. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID  
NO:109.  
30

165. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID  
NO:110.

166. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:111.

167. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID  
5 NO:112.

168. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:113.

10 169. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:114.

<110> INCYTE GENOMICS, INC.

AZIMZAI, Yalda  
BAUGHN, Mariah R.  
BECHA, Shanya D.  
BOROWSKY, Mark L.  
BURFORD, Neil  
ELLIOTT, Vicki S.  
EMERLING, Brooke M.  
FORSYTHE, Ian J.  
GORVAD, Ann E.  
GRIFFIN, Jennifer A.  
KABLE, Amy E.  
KHARE, Reena  
LAL, Preeti G.  
LEE, Ernestine A.  
LEE, Soo Yeun  
LI, Joana X.  
MARQUIS, Joseph P.  
RAMKUMAR, Jayalaxmi  
RICHARDSON, Thomas W.  
SPRAGUE, William W.  
SWARNAKER, Anita  
TANG, Y. Tom  
CHAWLA, Narinder K.  
WARREN, Bridget A.  
YUE, Henry

<120> NUCLEIC ACID-ASSOCIATED PROTEINS

<130> PF-1338 PCT

<140> To Be Assigned

<141> Herewith

<150> US 60/343,004

<151> 2001-12-19

<150> US 60/347,633

<151> 2002-01-11

<150> US 60/351,749

<151> 2002-01-25

<150> US 60/359,498

<151> 2002-02-22

<160> 114

<170> PERL Program

<210> 1

<211> 709

<212> PRT

<213> Homo sapiens

<220>

<221> misc\_feature

&lt;223&gt; Incyte ID No: 7506140CD1

&lt;400&gt; 1

Met	Ala	Ala	Asp	Ser	Arg	Glu	Glu	Lys	Asp	Gly	Glu	Leu	Asn	Val
1				5					10					15
Leu	Asp	Asp	Ile	Leu	Thr	Glu	Val	Pro	Glu	Gln	Asp	Asp	Glu	Leu
				20					25					30
Tyr	Asn	Pro	Glu	Ser	Glu	Gln	Asp	Lys	Asn	Glu	Lys	Lys	Gly	Ser
				35					40					45
Lys	Arg	Lys	Ser	Asp	Arg	Met	Glu	Ser	Thr	Asp	Thr	Lys	Arg	Gln
				50					55					60
Lys	Pro	Ser	Val	His	Ser	Arg	Gln	Leu	Val	Ser	Lys	Pro	Leu	Ser
				65					70					75
Ser	Ser	Val	Ser	Asn	Asn	Lys	Arg	Ile	Val	Ser	Thr	Lys	Gly	Lys
				80					85					90
Ser	Ala	Thr	Glu	Tyr	Lys	Asn	Glu	Glu	Tyr	Gln	Arg	Ser	Glu	Arg
				95					100					105
Asn	Lys	Arg	Leu	Asp	Ala	Asp	Arg	Lys	Ile	Arg	Leu	Ser	Ser	Ser
				110					115					120
Ala	Ser	Arg	Glu	Pro	Tyr	Lys	Asn	Gln	Pro	Glu	Lys	Thr	Cys	Val
				125					130					135
Arg	Lys	Arg	Asp	Pro	Glu	Arg	Arg	Ala	Lys	Ser	Pro	Thr	Pro	Asp
				140					145					150
Gly	Ser	Glu	Arg	Ile	Gly	Leu	Glu	Val	Asp	Arg	Arg	Ala	Ser	Arg
				155					160					165
Ser	Ser	Gln	Ser	Ser	Lys	Glu	Glu	Val	Asn	Ser	Glu	Glu	Tyr	Gly
				170					175					180
Ser	Asp	His	Glu	Thr	Gly	Ser	Ser	Gly	Ser	Ser	Asp	Glu	Gln	Gly
				185					190					195
Asn	Asn	Thr	Glu	Asn	Glu	Glu	Glu	Gly	Val	Glu	Glu	Asp	Val	Glu
				200					205					210
Glu	Asp	Glu	Glu	Glu	Glu	Glu	Glu	Glu	Glu	Glu	Glu	Glu	Tyr	Glu
				215					220					225
Gln	Asp	Glu	Arg	Asp	Gln	Lys	Glu	Glu	Gly	Asn	Asp	Tyr	Asp	Thr
				230					235					240
Arg	Ser	Glu	Ala	Ser	Asp	Ser	Gly	Ser	Glu	Ser	Val	Ser	Phe	Thr
				245					250					255
Asp	Gly	Ser	Val	Arg	Ser	Gly	Ser	Gly	Thr	Asp	Gly	Ser	Asp	Glu
				260					265					270
Lys	Lys	Lys	Glu	Arg	Lys	Arg	Ala	Arg	Gly	Ile	Ser	Pro	Ile	Val
				275					280					285
Phe	Asp	Arg	Ser	Gly	Ser	Ser	Ala	Ser	Glu	Ser	Tyr	Ala	Gly	Ser
				290					295					300
Glu	Lys	Lys	His	Glu	Lys	Leu	Ser	Ser	Ser	Val	Arg	Ala	Val	Arg
				305					310					315
Lys	Asp	Gln	Thr	Ser	Lys	Leu	Lys	Tyr	Val	Leu	Gln	Asp	Ala	Arg
				320					325					330
Phe	Phe	Leu	Ile	Lys	Ser	Asn	Asn	His	Glu	Asn	Val	Ser	Leu	Ala
				335					340					345
Lys	Ala	Lys	Gly	Val	Trp	Ser	Thr	Leu	Pro	Val	Asn	Glu	Lys	Lys
				350					355					360
Leu	Asn	Leu	Ala	Phe	Arg	Ser	Ala	Arg	Ser	Val	Ile	Leu	Ile	Phe
				365					370					375
Ser	Val	Arg	Glu	Ser	Gly	Lys	Phe	Gln	Gly	Phe	Ala	Arg	Leu	Ser
				380					385					390
Ser	Glu	Ser	His	His	Gly	Gly	Ser	Pro	Ile	His	Trp	Val	Leu	Pro

	395		400		405
Ala Gly Met Ser	Ala Lys Met Leu Gly Gly Val Phe Lys Ile Asp				
	410		415		420
Trp Ile Cys Arg	Arg Glu Leu Pro Phe Thr Lys Ser Ala His Leu				
	425		430		435
Thr Asn Pro Trp	Asn Glu His Lys Pro Val Lys Ile Gly Arg Asp				
	440		445		450
Gly Gln Glu Ile	Glu Leu Glu Cys Gly Thr Gln Leu Cys Leu Leu				
	455		460		465
Phe Pro Pro Asp	Glu Ser Ile Asp Leu Tyr Gln Val Ile His Lys				
	470		475		480
Met Arg His Lys	Arg Arg Met His Ser Gln Pro Arg Ser Arg Gly				
	485		490		495
Arg Pro Ser Arg	Arg Glu Pro Val Arg Asp Val Gly Arg Arg Arg				
	500		505		510
Pro Glu Asp Tyr	Asp Ile His Asn Ser Arg Lys Lys Pro Arg Ile				
	515		520		525
Asp Tyr Pro Pro	Glu Phe His Gln Arg Pro Gly Tyr Leu Lys Asp				
	530		535		540
Pro Arg Tyr Gln	Glu Val Asp Ser Phe Thr Asn Leu Ile Pro Asn				
	545		550		555
Arg Arg Phe Ser	Gly Val Arg Arg Asp Val Phe Leu Asn Gly Ser				
	560		565		570
Tyr Asn Asp Tyr	Val Arg Glu Phe His Asn Met Gly Pro Pro Pro				
	575		580		585
Pro Trp Gln Gly	Met Pro Pro Tyr Pro Gly Met Glu Gln Pro Pro				
	590		595		600
His His Pro Tyr	Tyr Gln His His Ala Pro Pro Pro Gln Ala His				
	605		610		615
Pro Pro Tyr Ser	Gly His His Pro Val Pro His Glu Ala Arg Tyr				
	620		625		630
Arg Asp Lys Arg	Val His Asp Tyr Asp Met Arg Val Asp Asp Phe				
	635		640		645
Leu Arg Arg Thr	Gln Ala Val Val Ser Gly Arg Arg Ser Arg Pro				
	650		655		660
Arg Glu Arg Asp	Arg Glu Arg Glu Arg Asp Arg Pro Arg Asp Asn				
	665		670		675
Arg Arg Asp Arg	Glu Arg Asp Arg Gly Arg Asp Arg Glu Arg Glu				
	680		685		690
Arg Glu Arg Leu	Cys Asp Arg Asp Arg Asp Arg Gly Glu Arg Gly				
	695		700		705
Arg Tyr Arg Arg					

&lt;210&gt; 2

&lt;211&gt; 986

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1889415CD1

&lt;400&gt; 2

Met Ser Phe Ala Leu Glu Glu Thr Leu Glu Ser Asp Trp Val Ala
1 5 10 15



Val	Arg	Pro	His	Val	Phe	Asp	Glu	Arg	Glu	Lys	His	Lys	Phe	Val
				20					25					30
Phe	Ile	Val	Ala	Trp	Asn	Glu	Ile	Glu	Gly	Lys	Phe	Ala	Ile	Thr
				35					40					45
Cys	His	Asn	Arg	Thr	Ala	Gln	Arg	Gln	Arg	Ser	Gly	Ser	Arg	Glu
				50					55					60
Gln	Val	Gly	Ala	Arg	Gly	Gly	Ala	Glu	Ala	Gly	Gly	Ala	Ala	Ser
				65					70					75
Asp	Gly	Ser	Arg	Gly	Pro	Gly	Ser	Pro	Ala	Gly	Arg	Gly	Arg	Pro
				80					85					90
Glu	Ala	Thr	Ala	Ser	Ala	Thr	Leu	Val	Arg	Ser	Pro	Gly	Pro	Arg
				95					100					105
Arg	Ser	Ser	Ala	Trp	Ala	Glu	Gly	Gly	Ser	Pro	Arg	Ser	Thr	Arg
				110					115					120
Ser	Leu	Leu	Gly	Asp	Pro	Arg	Leu	Arg	Ser	Pro	Gly	Ser	Lys	Gly
				125					130					135
Ala	Glu	Ser	Arg	Leu	Arg	Ser	Pro	Val	Arg	Ala	Lys	Pro	Ile	Pro
				140					145					150
Gly	Gln	Lys	Thr	Ser	Glu	Ala	Asp	Asp	Ala	Ala	Gly	Ala	Ala	Ala
				155					160					165
Ala	Ala	Ala	Arg	Pro	Ala	Pro	Arg	Glu	Ala	Gln	Val	Ser	Ser	Val
				170					175					180
Arg	Ile	Val	Ser	Ala	Ser	Gly	Thr	Val	Ser	Glu	Glu	Ile	Glu	Val
				185					190					195
Leu	Glu	Met	Val	Lys	Glu	Asp	Glu	Ala	Pro	Leu	Ala	Leu	Ser	Asp
				200					205					210
Ala	Glu	Gln	Pro	Pro	Pro	Ala	Thr	Glu	Leu	Glu	Ser	Pro	Ala	Glu
				215					220					225
Glu	Cys	Ser	Trp	Ala	Gly	Leu	Phe	Ser	Phe	Gln	Asp	Leu	Arg	Ala
				230					235					240
Val	His	Gln	Gln	Leu	Cys	Ser	Val	Asn	Ser	Gln	Leu	Glu	Pro	Cys
				245					250					255
Leu	Pro	Val	Phe	Pro	Glu	Glu	Pro	Ser	Gly	Met	Trp	Thr	Val	Leu
				260					265					270
Phe	Gly	Gly	Ala	Pro	Glu	Met	Thr	Glu	Gln	Glu	Ile	Asp	Thr	Leu
				275					280					285
Cys	Tyr	Gln	Leu	Gln	Val	Tyr	Leu	Gly	His	Gly	Leu	Asp	Thr	Cys
				290					295					300
Gly	Trp	Lys	Ile	Leu	Ser	Gln	Val	Leu	Phe	Thr	Glu	Thr	Asp	Asp
				305					310					315
Pro	Glu	Glu	Tyr	Tyr	Glu	Ser	Leu	Ser	Glu	Leu	Arg	Gln	Lys	Gly
				320					325					330
Tyr	Glu	Glu	Val	Leu	Gln	Arg	Ala	Arg	Lys	Arg	Ile	Gln	Glu	Leu
				335					340					345
Leu	Asp	Lys	His	Lys	Asn	Thr	Glu	Ser	Met	Val	Glu	Leu	Leu	Asp
				350					355					360
Leu	Tyr	Gln	Leu	Glu	Asp	Glu	Ala	Tyr	Ser	Ser	Leu	Ala	Glu	Ala
				365					370					375
Thr	Thr	Glu	Leu	Tyr	Gln	Tyr	Leu	Leu	Gln	Pro	Phe	Arg	Asp	Met
				380					385					390
Arg	Glu	Leu	Ala	Met	Leu	Arg	Arg	Gln	Gln	Ile	Lys	Ile	Ser	Met
				395					400					405
Glu	Asn	Asp	Tyr	Leu	Gly	Pro	Arg	Arg	Ile	Glu	Ser	Leu	Gln	Lys
				410					415					420
Glu	Asp	Ala	Asp	Trp	Gln	Arg	Lys	Ala	His	Met	Ala	Val	Leu	Ser
				425					430					435

Ile	Gln	Asp	Leu	Thr	Val	Lys	Tyr	Phe	Glu	Ile	Thr	Ala	Lys	Ala
				440					445					450
Gln	Lys	Ala	Val	Tyr	Asp	Arg	Met	Arg	Ala	Asp	Gln	Lys	Lys	Phe
				455					460					465
Gly	Lys	Ala	Ser	Trp	Ala	Ala	Ala	Ala	Glu	Arg	Met	Glu	Lys	Leu
				470					475					480
Gln	Tyr	Ala	Val	Ser	Lys	Glu	Thr	Leu	Gln	Met	Met	Arg	Ala	Lys
				485					490					495
Glu	Ile	Cys	Leu	Glu	Gln	Arg	Lys	His	Ala	Leu	Lys	Glu	Glu	Met
				500					505					510
Gln	Ser	Leu	Arg	Gly	Gly	Thr	Glu	Ala	Ile	Ala	Arg	Leu	Asp	Gln
				515					520					525
Leu	Glu	Ala	Asp	Tyr	Tyr	Asp	Leu	Gln	Leu	Gln	Leu	Tyr	Glu	Val
				530					535					540
Gln	Phe	Glu	Ile	Leu	Lys	Cys	Glu	Glu	Leu	Leu	Leu	Thr	Ala	Gln
				545					550					555
Leu	Glu	Ser	Ile	Lys	Arg	Leu	Ile	Ser	Glu	Lys	Arg	Asp	Glu	Val
				560					565					570
Val	Tyr	Tyr	Asp	Thr	Tyr	Glu	Ser	Met	Glu	Ala	Met	Leu	Glu	Lys
				575					580					585
Glu	Glu	Met	Ala	Ala	Ser	Ala	Tyr	Leu	Gln	Arg	Glu	Glu	Leu	Gln
				590					595					600
Lys	Leu	Gln	Gln	Lys	Ala	Arg	Gln	Leu	Glu	Ala	Arg	Arg	Gly	Arg
				605					610					615
Val	Ser	Ala	Lys	Lys	Ser	Tyr	Leu	Arg	Asn	Lys	Lys	Glu	Ile	Cys
				620					625					630
Ile	Ala	Lys	His	Asn	Glu	Ile	Gln	Gln	Arg	Thr	Arg	Ile	Glu	Asp
				635					640					645
Glu	Tyr	Arg	Thr	His	His	Thr	Val	Gln	Leu	Arg	Glu	Lys	Leu	His
				650					655					660
Asp	Glu	Glu	Glu	Arg	Lys	Ser	Ala	Trp	Val	Ser	Gln	Glu	Arg	Gln
				665					670					675
Arg	Thr	Leu	Asp	Arg	Leu	Arg	Thr	Phe	Lys	Gln	Arg	Tyr	Pro	Gly
				680					685					690
Gln	Val	Ile	Leu	Lys	Ser	Thr	Arg	Leu	Arg	Leu	Ala	His	Ala	Arg
				695					700					705
Arg	Lys	Gly	Ala	Ala	Ser	Pro	Val	Leu	Gln	Glu	Asp	His	Cys	Asp
				710					715					720
Ser	Leu	Pro	Ser	Val	Leu	Gln	Val	Glu	Glu	Lys	Thr	Glu	Glu	Val
				725					730					735
Gly	Glu	Gly	Arg	Val	Lys	Arg	Gly	Pro	Ser	Gln	Thr	Thr	Glu	Pro
				740					745					750
Gln	Ser	Leu	Val	Gln	Leu	Glu	Asp	Thr	Ser	Leu	Thr	Gln	Leu	Glu
				755					760					765
Ala	Thr	Ser	Leu	Pro	Leu	Ser	Gly	Val	Thr	Ser	Glu	Leu	Pro	Pro
				770					775					780
Thr	Ile	Ser	Leu	Pro	Leu	Leu	Asn	Asn	Asn	Leu	Glu	Pro	Cys	Ser
				785					790					795
Val	Thr	Ile	Asn	Pro	Leu	Pro	Ser	Pro	Leu	Pro	Pro	Thr	Pro	Pro
				800					805					810
Pro	Pro	Pro	Pro	Pro	Pro	Pro	Pro	Pro	Pro	Pro	Pro	Pro	Leu	Pro
				815					820					825
Val	Ala	Lys	Asp	Ser	Gly	Pro	Glu	Thr	Leu	Glu	Lys	Asp	Leu	Pro
				830					835					840
Arg	Lys	Glu	Gly	Asn	Asp	Lys	Arg	Ile	Pro	Lys	Ser	Ala	Ser	Ala
				845					850					855

```

Pro Ser Ala His Leu Phe Asp Ser Ser Gln Leu Val Ser Ala Arg
      860      865      870
Lys Lys Leu Arg Lys Thr Ala Glu Gly Leu Gln Arg Arg Arg Val
      875      880      885
Ser Ser Pro Met Asp Glu Val Leu Ala Ser Leu Lys Arg Gly Ser
      890      895      900
Phe His Leu Lys Lys Val Glu Gln Arg Thr Leu Pro Pro Phe Pro
      905      910      915
Asp Glu Asp Asp Ser Asn Asn Ile Leu Ala Gln Ile Arg Lys Gly
      920      925      930
Val Lys Leu Lys Lys Val Gln Lys Asp Val Leu Arg Glu Ser Phe
      935      940      945
Thr Leu Leu Pro Asp Thr Asp Pro Leu Thr Arg Ser Ile His Glu
      950      955      960
Ala Leu Arg Arg Ile Lys Glu Ala Ser Pro Glu Ser Glu Asp Glu
      965      970      975
Glu Glu Ala Leu Pro Cys Thr Asp Trp Glu Asn
      980      985

```

```

<210> 3
<211> 714
<212> PRT
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7506047CD1

```

```

<400> 3
Met Pro Asp His Asp Ser Thr Ala Leu Leu Ser Arg Gln Thr Lys
  1          5          10          15
Arg Arg Arg Val Asp Ile Gly Val Lys Arg Thr Val Gly Thr Ala
      20      25      30
Ser Ala Phe Phe Ala Lys Ala Arg Ala Thr Phe Phe Ser Ala Met
      35      40      45
Asn Pro Gln Gly Ser Glu Gln Asp Val Glu Tyr Ser Val Val Gln
      50      55      60
His Ala Asp Gly Glu Lys Ser Asn Val Leu Arg Lys Leu Leu Lys
      65      70      75
Asn Asn Met Asn Lys Asn Gly Gly Thr Glu Pro Ser Phe Gln Ala
      80      85      90
Ser Gly Leu Ser Ser Thr Gly Ser Glu Val His Gln Glu Asp Ile
      95      100      105
Cys Ser Asn Ser Ser Arg Asp Ser Pro Pro Glu Cys Leu Ser Pro
      110      115      120
Phe Gly Arg Pro Thr Met Ser Gln Phe Asp Met Asp Arg Leu Cys
      125      130      135
Asp Glu His Leu Arg Ala Lys Arg Ala Arg Val Glu Asn Ile Ile
      140      145      150
Arg Gly Met Ser His Ser Pro Ser Val Ala Leu Arg Gly Asn Glu
      155      160      165
Asn Glu Arg Glu Met Ala Pro Gln Ser Val Ser Pro Arg Glu Ser
      170      175      180
Tyr Arg Glu Asn Lys Arg Lys Gln Lys Leu Pro Gln Gln Gln Gln
      185      190      195
Gln Ser Phe Gln Gln Leu Val Ser Ala Arg Lys Glu Gln Lys Arg

```

	200	205	210
Glu Glu Arg Arg	Gln Leu Lys Gln Gln	Leu Glu Asp Met Gln Lys	
	215	220	225
Gln Leu Arg Gln	Leu Gln Glu Lys Phe	Tyr Gln Ile Tyr Asp Ser	
	230	235	240
Thr Asp Ser Glu	Asn Asp Glu Asp Gly	Asn Leu Ser Glu Asp Ser	
	245	250	255
Met Arg Ser Glu	Ile Leu Asp Ala Arg	Gln Asp Ser Val Gly	
	260	265	270
Arg Ser Asp Asn	Glu Met Cys Glu Leu	Asp Pro Gly Gln Phe Ile	
	275	280	285
Asp Arg Ala Arg	Ala Leu Ile Arg Glu	Gln Glu Met Ala Glu Asn	
	290	295	300
Lys Pro Lys Arg	Glu Gly Asn Asn Lys	Glu Arg Asp His Gly Pro	
	305	310	315
Asn Ser Leu Gln	Pro Glu Gly Lys His	Leu Ala Glu Thr Leu Lys	
	320	325	330
Gln Glu Leu Asn	Thr Ala Met Ser Gln	Val Val Asp Thr Val Val	
	335	340	345
Lys Val Phe Ser	Ala Lys Pro Ser Arg	Gln Val Pro Gln Val Phe	
	350	355	360
Pro Pro Leu Gln	Ile Pro Gln Ala Arg	Phe Ala Val Asn Gly Glu	
	365	370	375
Asn His Asn Phe	His Thr Ala Asn Gln	Arg Leu Gln Cys Phe Gly	
	380	385	390
Asp Val Ile Ile	Pro Asn Pro Leu Asp	Thr Phe Gly Asn Val Gln	
	395	400	405
Met Ala Ser Ser	Thr Asp Gln Thr Glu	Ala Leu Pro Leu Val Val	
	410	415	420
Arg Lys Asn Ser	Ser Asp Gln Ser Ala	Ser Gly Pro Ala Ala Gly	
	425	430	435
Gly His His Gln	Pro Leu His Gln Ser	Pro Leu Ser Ala Thr Thr	
	440	445	450
Gly Phe Thr Thr	Ser Thr Phe Arg His	Pro Phe Pro Leu Pro Leu	
	455	460	465
Met Ala Tyr Pro	Phe Gln Ser Pro Leu	Gly Ala Pro Ser Gly Ser	
	470	475	480
Phe Ser Gly Lys	Asp Arg Ala Ser Pro	Glu Ser Leu Asp Leu Thr	
	485	490	495
Arg Asp Thr Thr	Ser Leu Arg Thr Lys	Met Ser Ser His His Leu	
	500	505	510
Ser His His Pro	Cys Ser Pro Ala His	Pro Pro Ser Thr Ala Glu	
	515	520	525
Gly Leu Ser Leu	Ser Leu Ile Lys Ser	Glu Cys Gly Asp Leu Gln	
	530	535	540
Asp Met Ser Glu	Ile Ser Pro Tyr Ser	Gly Ser Ala Met Gln Glu	
	545	550	555
Gly Leu Ser Pro	Asn His Leu Lys Lys	Ala Lys Leu Met Phe Phe	
	560	565	570
Tyr Thr Arg Tyr	Pro Ser Ser Asn Met	Leu Lys Thr Tyr Phe Ser	
	575	580	585
Asp Val Lys Phe	Asn Arg Cys Ile Thr	Ser Gln Leu Ile Lys Trp	
	590	595	600
Phe Ser Asn Phe	Arg Glu Phe Tyr Tyr	Ile Gln Met Glu Lys Tyr	
	605	610	615
Ala Arg Gln Ala	Ile Asn Asp Gly Val	Thr Ser Thr Glu Glu Leu	

	620		625		630
Ser Ile Thr Arg Asp Cys Glu Leu Tyr		Arg Ala Leu Asn Met His			
	635		640		645
Tyr Asn Lys Ala Asn Asp Phe Glu Val		Pro Glu Arg Phe Leu Glu			
	650		655		660
Val Ala Gln Ile Thr Leu Arg Glu Phe		Phe Asn Ala Ile Ile Ala			
	665		670		675
Gly Lys Asp Val Asp Pro Ser Trp Lys		Lys Ala Ile Tyr Lys Val			
	680		685		690
Ile Cys Lys Leu Asp Ser Glu Val Pro		Glu Ile Phe Lys Ser Pro			
	695		700		705
Asn Cys Leu Gln Glu Leu Leu His Glu					
	710				

&lt;210&gt; 4

&lt;211&gt; 166

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7505849CD1

&lt;400&gt; 4

Met Ser Lys Leu Gly Arg Ala Ala Arg Gly Leu Arg Lys Pro Glu		
1	5	10 15
Arg Gly Val Ser Ile Asn Gln Phe Cys Lys Glu Phe Asn Glu Arg		
	20	25 30
Thr Lys Asp Ile Lys Glu Gly Ile Pro Leu Pro Thr Lys Ile Leu		
	35	40 45
Val Lys Pro Asp Arg Thr Phe Glu Ile Lys Ile Gly Gln Pro Thr		
	50	55 60
Val Ser Tyr Phe Leu Lys Ala Ala Ala Gly Ile Glu Lys Gly Ala		
	65	70 75
Arg Gln Thr Gly Lys Glu Val Ala Gly Leu Val Thr Leu Lys His		
	80	85 90
Val Tyr Glu Ile Ala Arg Ile Lys Ala Gln Asp Glu Ala Phe Ala		
	95	100 105
Leu Gln Asp Val Pro Leu Ser Ser Val Val Arg Ser Ile Ile Gly		
	110	115 120
Ser Ala Arg Ser Leu Gly Ile Arg Val Val Lys Asp Leu Ser Ser		
	125	130 135
Glu Glu Leu Ala Ala Phe Gln Lys Glu Arg Ala Ile Phe Leu Ala		
	140	145 150
Ala Gln Lys Glu Ala Asp Leu Ala Ala Gln Glu Glu Ala Ala Lys		
	155	160 165
Lys		

&lt;210&gt; 5

&lt;211&gt; 142

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7505972CD1

&lt;400&gt; 5

Met	Ala	Ser	Asp	Glu	Gly	Lys	Leu	Phe	Val	Gly	Gly	Leu	Ser	Phe
1				5					10					15
Asp	Thr	Asn	Glu	Gln	Ser	Leu	Glu	Gln	Val	Phe	Ser	Lys	Tyr	Gly
			20						25					30
Gln	Ile	Ser	Glu	Val	Val	Val	Lys	Asp	Arg	Glu	Thr	Gln	Arg	
			35						40					45
Ser	Arg	Gly	Phe	Gly	Phe	Val	Thr	Phe	Glu	Asn	Ile	Asp	Asp	Ala
			50						55					60
Lys	Asp	Ala	Met	Met	Ala	Met	Asn	Gly	Lys	Ser	Val	Asp	Gly	Arg
			65						70					75
Gln	Ile	Arg	Gly	Arg	Gly	Arg	Gly	Phe	Ser	Arg	Gly	Gly	Gly	Asp
			80						85					90
Arg	Gly	Tyr	Gly	Gly	Asn	Arg	Phe	Glu	Ser	Arg	Ser	Gly	Gly	Tyr
			95						100					105
Gly	Gly	Ser	Arg	Asp	Tyr	Tyr	Ser	Ser	Arg	Ser	Gln	Ser	Gly	Gly
			110						115					120
Tyr	Ser	Asp	Arg	Ser	Ser	Gly	Gly	Ser	Tyr	Arg	Asp	Ser	Tyr	Asp
			125						130					135
Ser	Tyr	Ala	Thr	His	Asn	Glu								
														140

&lt;210&gt; 6

&lt;211&gt; 317

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7505991CD1

&lt;400&gt; 6

Met	Glu	Val	Leu	Arg	Pro	Gln	Leu	Ile	Arg	Ile	Asp	Gly	Arg	Asn
1				5					10					15
Tyr	Arg	Lys	Asn	Pro	Val	Gln	Glu	Gln	Thr	Tyr	Gln	His	Glu	Glu
			20						25					30
Asp	Glu	Glu	Asp	Phe	Tyr	Gln	Gly	Ser	Met	Glu	Cys	Ala	Asp	Glu
			35						40					45
Pro	Cys	Asp	Ala	Tyr	Glu	Val	Glu	Gln	Thr	Pro	Gln	Gly	Phe	Arg
			50						55					60
Ser	Thr	Leu	Arg	Ala	Pro	Ser	Leu	Leu	Tyr	Lys	His	Ile	Val	Gly
			65						70					75
Lys	Arg	Gly	Asp	Thr	Arg	Lys	Lys	Ile	Glu	Met	Glu	Thr	Lys	Thr
			80						85					90
Ser	Ile	Ser	Ile	Pro	Lys	Pro	Gly	Gln	Asp	Gly	Glu	Ile	Val	Ile
			95						100					105
Thr	Gly	Gln	His	Arg	Asn	Gly	Val	Ile	Ser	Ala	Arg	Thr	Arg	Ile
			110						115					120
Asp	Val	Leu	Leu	Asp	Thr	Phe	Arg	Arg	Lys	Gln	Pro	Phe	Thr	His
			125						130					135
Phe	Leu	Ala	Phe	Phe	Leu	Asn	Glu	Val	Glu	Val	Gln	Glu	Gly	Phe
			140						145					150
Leu	Arg	Phe	Gln	Glu	Glu	Val	Leu	Ala	Lys	Cys	Ser	Met	Asp	His
			155						160					165

Gly	Val	Asp	Ser	Ser	Ile	Phe	Gln	Asn	Pro	Lys	Lys	Leu	His	Leu
				170					175					180
Thr	Ile	Gly	Met	Leu	Val	Leu	Leu	Ser	Glu	Glu	Glu	Ile	Gln	Gln
				185					190					195
Thr	Cys	Glu	Met	Leu	Gln	Gln	Cys	Lys	Glu	Glu	Phe	Ile	Lys	Leu
				200					205					210
Gln	Glu	Leu	Val	Asp	Arg	Val	Leu	Glu	Arg	Phe	Gln	Ala	Ser	Gly
				215					220					225
Leu	Ile	Val	Lys	Glu	Trp	Asn	Ser	Val	Lys	Leu	His	Ala	Thr	Val
				230					235					240
Met	Asn	Thr	Leu	Phe	Arg	Lys	Asp	Pro	Asn	Ala	Glu	Gly	Arg	Tyr
				245					250					255
Asn	Leu	Tyr	Thr	Ala	Glu	Gly	Lys	Tyr	Ile	Phe	Lys	Glu	Arg	Glu
				260					265					270
Ser	Phe	Asp	Gly	Arg	Asn	Ile	Leu	Lys	Leu	Phe	Glu	Asn	Phe	Tyr
				275					280					285
Phe	Gly	Ser	Leu	Lys	Leu	Asn	Ser	Ile	His	Ile	Ser	Gln	Arg	Phe
				290					295					300
Thr	Val	Asp	Ser	Phe	Gly	Asn	Tyr	Ala	Ser	Cys	Gly	Gln	Ile	Asp
				305					310					315

Phe Ser

&lt;210&gt; 7

&lt;211&gt; 1359

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7506003CD1

&lt;400&gt; 7

Met	Val	Lys	Glu	Gln	Phe	Arg	Glu	Thr	Asp	Val	Ala	Lys	Lys	Ile
1				5					10					15
Ser	His	Ile	Cys	Phe	Gly	Met	Lys	Ser	Pro	Glu	Glu	Met	Arg	Gln
				20					25					30
Gln	Ala	His	Ile	Gln	Val	Val	Ser	Lys	Asn	Leu	Tyr	Ser	Gln	Asp
				35					40					45
Asn	Gln	His	Ala	Pro	Leu	Leu	Tyr	Gly	Val	Leu	Asp	His	Arg	Met
				50					55					60
Gly	Thr	Ser	Glu	Lys	Asp	Arg	Pro	Cys	Glu	Thr	Cys	Gly	Lys	Asn
				65					70					75
Leu	Ala	Asp	Cys	Leu	Gly	His	Tyr	Gly	Tyr	Ile	Asp	Leu	Glu	Leu
				80					85					90
Pro	Cys	Phe	His	Val	Gly	Tyr	Phe	Arg	Ala	Val	Ile	Gly	Ile	Leu
				95					100					105
Gln	Met	Ile	Cys	Lys	Thr	Cys	Cys	His	Ile	Met	Leu	Ser	Gln	Glu
				110					115					120
Glu	Lys	Lys	Gln	Phe	Leu	Asp	Tyr	Leu	Lys	Arg	Pro	Gly	Leu	Thr
				125					130					135
Tyr	Leu	Gln	Lys	Arg	Gly	Leu	Lys	Lys	Lys	Ile	Ser	Asp	Lys	Cys
				140					145					150
Arg	Lys	Lys	Asn	Ile	Cys	His	His	Cys	Gly	Ala	Phe	Asn	Gly	Thr
				155					160					165
Val	Lys	Lys	Cys	Gly	Leu	Leu	Lys	Ile	Ile	His	Glu	Lys	Tyr	Lys

	170		175		180
Thr Asn Lys Lys Val Val Asp Pro Ile Val Ser Asn Phe Leu Gln					
	185		190		195
Ser Phe Glu Thr Ala Ile Glu His Asn Lys Glu Val Glu Pro Leu					
	200		205		210
Leu Gly Arg Ala Gln Glu Asn Leu Asn Pro Leu Val Val Leu Asn					
	215		220		225
Leu Phe Lys Arg Ile Pro Ala Glu Asp Val Pro Leu Leu Leu Met					
	230		235		240
Asn Pro Glu Ala Gly Lys Pro Ser Asp Leu Ile Leu Thr Arg Leu					
	245		250		255
Leu Val Pro Pro Leu Cys Phe Arg Pro Ser Val Val Ser Asp Leu					
	260		265		270
Lys Ser Gly Thr Asn Glu Asp Asp Leu Thr Met Lys Leu Thr Glu					
	275		280		285
Ile Ile Phe Leu Asn Asp Val Ile Lys Lys His Arg Ile Ser Gly					
	290		295		300
Ala Lys Thr Gln Met Ile Met Glu Asp Trp Asp Phe Leu Gln Leu					
	305		310		315
Gln Cys Ala Leu Tyr Ile Asn Ser Glu Leu Ser Gly Ile Pro Leu					
	320		325		330
Asn Met Ala Pro Lys Lys Trp Thr Arg Gly Phe Val Gln Arg Leu					
	335		340		345
Lys Gly Lys Gln Gly Arg Phe Arg Gly Asn Leu Ser Gly Lys Arg					
	350		355		360
Val Asp Phe Ser Gly Arg Thr Val Ile Ser Pro Asp Pro Asn Leu					
	365		370		375
Arg Ile Asp Glu Val Ala Val Pro Val His Val Ala Lys Ile Leu					
	380		385		390
Thr Phe Pro Glu Lys Val Asn Lys Ala Asn Ile Asn Phe Leu Arg					
	395		400		405
Lys Leu Val Gln Asn Gly Pro Glu Val His Pro Gly Ala Asn Phe					
	410		415		420
Ile Gln Gln Arg His Thr Gln Met Lys Arg Phe Leu Lys Tyr Gly					
	425		430		435
Asn Arg Glu Lys Met Ala Gln Glu Leu Lys Tyr Gly Asp Ile Val					
	440		445		450
Glu Arg His Leu Ile Asp Gly Asp Val Val Leu Phe Asn Arg Gln					
	455		460		465
Pro Ser Leu His Lys Leu Ser Ile Met Ala His Leu Ala Arg Val					
	470		475		480
Lys Pro His Arg Thr Phe Arg Phe Asn Glu Cys Val Cys Thr Pro					
	485		490		495
Tyr Asn Ala Asp Phe Asp Gly Asp Glu Met Asn Leu His Leu Pro					
	500		505		510
Gln Thr Glu Glu Ala Lys Ala Glu Ala Leu Val Leu Met Gly Thr					
	515		520		525
Lys Ala Asn Leu Val Thr Pro Arg Asn Gly Glu Pro Leu Ile Ala					
	530		535		540
Ala Ile Gln Asp Phe Leu Thr Gly Ala Tyr Leu Leu Thr Leu Lys					
	545		550		555
Asp Thr Phe Phe Asp Arg Ala Lys Ala Cys Gln Ile Ile Ala Ser					
	560		565		570
Ile Leu Val Gly Lys Asp Glu Lys Ile Lys Val Arg Leu Pro Pro					
	575		580		585
Pro Thr Ile Leu Lys Pro Val Thr Leu Trp Thr Gly Lys Gln Ile					



				590					595				600	
Phe	Ser	Val	Ile	Leu	Arg	Pro	Ser	Asp	Asp	Asn	Pro	Val	Arg	Ala
				605					610					615
Asn	Leu	Arg	Thr	Lys	Gly	Lys	Gln	Tyr	Cys	Gly	Lys	Gly	Glu	Asp
				620					625					630
Leu	Cys	Ala	Asn	Asp	Ser	Tyr	Val	Thr	Ile	Gln	Asn	Ser	Glu	Leu
				635					640					645
Met	Ser	Gly	Ser	Met	Asp	Lys	Gly	Thr	Leu	Gly	Ser	Gly	Ser	Lys
				650					655					660
Asn	Asn	Ile	Phe	Tyr	Ile	Leu	Leu	Arg	Asp	Trp	Gly	Gln	Asn	Glu
				665					670					675
Ala	Ala	Asp	Ala	Met	Ser	Arg	Leu	Ala	Arg	Leu	Ala	Pro	Val	Tyr
				680					685					690
Leu	Ser	Asn	Arg	Gly	Phe	Ser	Ile	Gly	Ile	Gly	Asp	Val	Thr	Pro
				695					700					705
Gly	Gln	Gly	Leu	Leu	Lys	Ala	Lys	Tyr	Glu	Leu	Leu	Asn	Ala	Gly
				710					715					720
Tyr	Lys	Lys	Cys	Asp	Glu	Tyr	Ile	Glu	Ala	Leu	Asn	Thr	Gly	Lys
				725					730					735
Leu	Gln	Gln	Gln	Pro	Gly	Cys	Thr	Ala	Glu	Glu	Thr	Leu	Glu	Ala
				740					745					750
Leu	Ile	Leu	Lys	Glu	Leu	Ser	Val	Ile	Arg	Asp	His	Ala	Gly	Ser
				755					760					765
Ala	Cys	Leu	Arg	Glu	Leu	Asp	Lys	Ser	Asn	Ser	Pro	Leu	Thr	Met
				770					775					780
Ala	Leu	Cys	Gly	Ser	Lys	Gly	Ser	Phe	Ile	Asn	Ile	Ser	Gln	Met
				785					790					795
Ile	Ala	Cys	Val	Gly	Gln	Gln	Ala	Ile	Ser	Gly	Ser	Arg	Val	Pro
				800					805					810
Asp	Gly	Phe	Glu	Asn	Arg	Ser	Leu	Pro	His	Phe	Glu	Lys	His	Ser
				815					820					825
Lys	Leu	Pro	Ala	Ala	Lys	Gly	Phe	Val	Ala	Asn	Ser	Phe	Tyr	Ser
				830					835					840
Gly	Leu	Thr	Pro	Thr	Glu	Phe	Phe	Phe	His	Thr	Met	Ala	Gly	Arg
				845					850					855
Glu	Gly	Leu	Val	Asp	Thr	Ala	Val	Lys	Thr	Ala	Glu	Thr	Gly	Tyr
				860					865					870
Met	Gln	Arg	Arg	Leu	Val	Lys	Ser	Leu	Glu	Asp	Leu	Cys	Ser	Gln
				875					880					885
Tyr	Asp	Leu	Thr	Val	Arg	Ser	Ser	Thr	Gly	Asp	Ile	Ile	Gln	Phe
				890					895					900
Ile	Tyr	Gly	Gly	Asp	Gly	Leu	Asp	Pro	Ala	Ala	Met	Glu	Gly	Lys
				905					910					915
Asp	Glu	Pro	Leu	Glu	Phe	Lys	Arg	Val	Leu	Asp	Asn	Ile	Lys	Ala
				920					925					930
Val	Phe	Pro	Cys	Pro	Ser	Glu	Pro	Ala	Leu	Ser	Lys	Asn	Glu	Leu
				935					940					945
Ile	Leu	Thr	Thr	Glu	Ser	Ile	Met	Lys	Lys	Ser	Glu	Phe	Leu	Cys
				950					955					960
Cys	Gln	Asp	Ser	Phe	Leu	Gln	Glu	Ile	Lys	Lys	Phe	Ile	Lys	Gly
				965					970					975
Val	Ser	Glu	Lys	Ile	Lys	Lys	Thr	Arg	Asp	Lys	Tyr	Gly	Ile	Asn
				980					985					990
Asp	Asn	Gly	Thr	Thr	Glu	Pro	Arg	Val	Leu	Tyr	Gln	Leu	Asp	Arg
				995					1000					1005
Ile	Thr	Pro	Thr	Gln	Val	Glu	Lys	Phe	Leu	Glu	Thr	Cys	Arg	Asp

1010	1015	1020
Lys Tyr Met Arg Ala Gln Met Glu Pro Gly Ser Ala Val Gly Ala		
1025	1030	1035
Leu Cys Ala Gln Ser Ile Gly Glu Pro Gly Thr Gln Met Thr Leu		
1040	1045	1050
Lys Thr Phe His Phe Ala Gly Val Ala Ser Met Asn Ile Thr Leu		
1055	1060	1065
Gly Val Pro Arg Ile Lys Glu Ile Ile Asn Ala Ser Lys Ala Ile		
1070	1075	1080
Ser Thr Pro Ile Ile Thr Ala Gln Leu Asp Lys Asp Asp Asp Ala		
1085	1090	1095
Asp Tyr Ala Arg Leu Val Lys Gly Arg Ile Glu Lys Thr Leu Leu		
1100	1105	1110
Gly Glu Val Asn Ala Glu Thr Val Arg Tyr Ser Ile Cys Thr Ser		
1115	1120	1125
Lys Leu Arg Val Lys Pro Gly Asp Val Ala Val His Gly Glu Ala		
1130	1135	1140
Val Val Cys Val Thr Pro Arg Glu Asn Ser Lys Ser Ser Met Tyr		
1145	1150	1155
Tyr Val Leu Gln Phe Leu Lys Glu Asp Leu Pro Lys Val Val Val		
1160	1165	1170
Gln Gly Ile Pro Glu Val Ser Arg Ala Val Ile His Ile Asp Glu		
1175	1180	1185
Gln Ser Gly Lys Glu Lys Tyr Lys Leu Leu Val Glu Gly Asp Asn		
1190	1195	1200
Leu Arg Ala Val Met Ala Thr His Gly Val Lys Gly Thr Arg Thr		
1205	1210	1215
Thr Ser Asn Asn Thr Tyr Glu Val Glu Lys Thr Leu Gly Ile Glu		
1220	1225	1230
Ala Ala Arg Thr Thr Ile Ile Asn Glu Ile Gln Tyr Thr Met Val		
1235	1240	1245
Asn His Gly Met Ser Ile Asp Arg Arg His Val Met Leu Leu Ser		
1250	1255	1260
Asp Leu Met Thr Tyr Lys Gly Glu Val Leu Gly Ile Thr Arg Phe		
1265	1270	1275
Gly Leu Ala Lys Met Lys Glu Ser Val Leu Met Leu Ala Ser Phe		
1280	1285	1290
Glu Lys Thr Ala Asp His Leu Phe Asp Ala Ala Tyr Phe Gly Gln		
1295	1300	1305
Lys Asp Ser Val Cys Gly Val Ser Glu Cys Ile Ile Met Gly Ile		
1310	1315	1320
Pro Met Asn Ile Gly Thr Gly Leu Phe Lys Leu Leu His Lys Ala		
1325	1330	1335
Asp Arg Asp Pro Asn Pro Pro Lys Arg Pro Leu Ile Phe Asp Thr		
1340	1345	1350
Asn Glu Phe His Ile Pro Leu Val Thr		
1355		

&lt;210&gt; 8

&lt;211&gt; 226

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 6483977CD1

&lt;400&gt; 8

```

Met Ser Arg Tyr Gly Arg Tyr Gly Gly Glu Thr Lys Val Tyr Val
 1          5          10          15
Gly Asn Leu Gly Thr Gly Ala Gly Lys Gly Glu Leu Glu Arg Ala
          20          25          30
Phe Ser Tyr Tyr Gly Pro Leu Arg Thr Val Trp Ile Ala Arg Asn
          35          40          45
Pro Pro Gly Phe Ala Phe Val Glu Phe Glu Asp Pro Arg Asp Ala
          50          55          60
Glu Asp Ala Val Arg Gly Leu Asp Gly Lys Val Ile Cys Gly Ser
          65          70          75
Arg Val Arg Val Glu Leu Ser Thr Gly Met Pro Arg Arg Ser Arg
          80          85          90
Phe Asp Arg Pro Pro Ala Arg Arg Pro Phe Asp Pro Asn Asp Arg
          95          100          105
Cys Tyr Glu Cys Gly Glu Lys Gly His Tyr Ala Tyr Asp Cys His
          110          115          120
Arg Tyr Ser Arg Arg Arg Arg Ser Arg Ser Arg Ser Arg Ser His
          125          130          135
Ser Arg Ser Arg Gly Arg Arg Tyr Ser Arg Ser Arg Ser Arg Ser
          140          145          150
Arg Gly Arg Arg Ser Arg Ser Ala Ser Pro Arg Arg Ser Arg Ser
          155          160          165
Ile Ser Leu Arg Arg Ser Arg Ser Ala Ser Leu Arg Arg Ser Arg
          170          175          180
Ser Gly Ser Ile Lys Gly Ser Arg Tyr Phe Gln Ser Pro Ser Arg
          185          190          195
Ser Arg Ser Arg Ser Arg Ser Ile Ser Arg Pro Arg Ser Ser Arg
          200          205          210
Ser Pro Ser Gly Ser Pro Arg Arg Ser Ala Ser Pro Glu Arg Met
          215          220          225
Asp

```

&lt;210&gt; 9

&lt;211&gt; 76

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 6301777CD1

&lt;400&gt; 9

```

Met Gly Leu Pro Arg Arg Ala Gly Asp Ala Ala Glu Leu Arg Lys
 1          5          10          15
Ser Leu Lys Pro Leu Leu Glu Lys Arg Arg Arg Ala Arg Ile Asn
          20          25          30
Gln Ser Leu Ser Gln Leu Lys Gly Leu Ile Leu Pro Leu Leu Gly
          35          40          45
Arg Glu Asp Ala Ser Gly Trp His Thr Trp Leu Pro Leu His Ala
          50          55          60
Gln Asn Cys Leu Leu Leu Tyr Ile Gln Ala Pro Glu Gln Pro Pro
          65          70          75
Ala

```

<210> 10  
<211> 124  
<212> PRT  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 7505976CD1

<400> 10  
Met Ala Ala Thr Glu Pro Ile Leu Ala Ala Thr Gly Ser Pro Ala  
1 5 10 15  
Ala Val Pro Pro Glu Lys Leu Glu Gly Ala Gly Ser Ser Ser Ala  
20 25 30  
Pro Glu Arg Asn Cys Val Gly Ser Ser Leu Pro Glu Ala Ser Pro  
35 40 45  
Pro Ala Pro Glu Pro Ser Ser Pro Asn Ala Ala Val Pro Glu Ala  
50 55 60  
Ile Pro Thr Pro Arg Ala Ala Ala Ser Ala Ala Leu Glu Leu Pro  
65 70 75  
Leu Gly Pro Ala Pro Val Ser Val Ala Pro Gln Ala Glu Ala Glu  
80 85 90  
Ala Arg Ser Thr Pro Gly Pro Ala Gly Ser Arg Leu Gly Pro Glu  
95 100 105  
Thr Phe Arg Gln Arg Phe Arg Gln Phe Arg Arg Arg Thr Asp Val  
110 115 120  
Arg Ile Thr Gly

<210> 11  
<211> 488  
<212> PRT  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 7506016CD1

<400> 11  
Met Asn Gln Ser Ile Pro Val Ala Pro Thr Pro Pro Arg Arg Val  
1 5 10 15  
Arg Leu Lys Pro Trp Leu Val Ala Gln Val Asn Ser Cys Gln Tyr  
20 25 30  
Pro Gly Leu Gln Trp Val Asn Gly Glu Lys Lys Leu Phe Cys Ile  
35 40 45  
Pro Trp Arg His Ala Thr Arg His Gly Pro Ser Gln Asp Gly Asp  
50 55 60  
Asn Thr Ile Phe Lys Ala Trp Ala Lys Glu Thr Gly Lys Tyr Thr  
65 70 75  
Glu Gly Val Asp Glu Ala Asp Pro Ala Lys Trp Lys Ala Asn Leu  
80 85 90  
Arg Cys Ala Leu Asn Lys Ser Arg Asp Phe Arg Leu Ile Tyr Asp  
95 100 105  
Gly Pro Arg Asp Met Pro Pro Gln Pro Tyr Lys Ile Tyr Glu Val  
110 115 120  
Cys Ser Asn Gly Pro Ala Pro Thr Asp Ser Gln Pro Pro Glu Asp

	125		130		135
Tyr Ser Phe Gly	Ala Gly Glu Glu Glu Glu Glu Glu Glu Glu Leu				
	140		145		150
Gln Arg Met Leu	Pro Ser Leu Ser Leu Thr Glu Asp Val Lys Trp				
	155		160		165
Pro Pro Thr Leu	Gln Pro Pro Thr Leu Gln Pro Pro Val Val Leu				
	170		175		180
Gly Pro Pro Ala	Pro Asp Pro Ser Pro Leu Ala Pro Pro Pro Gly				
	185		190		195
Asn Pro Ala Gly	Phe Arg Glu Leu Leu Ser Glu Val Leu Glu Pro				
	200		205		210
Gly Pro Leu Pro	Ala Ser Leu Pro Pro Ala Gly Glu Gln Leu Leu				
	215		220		225
Pro Asp Leu Leu	Ile Ser Pro His Met Leu Pro Leu Thr Asp Leu				
	230		235		240
Glu Ile Lys Phe	Gln Tyr Arg Gly Arg Pro Pro Arg Ala Leu Thr				
	245		250		255
Ile Ser Asn Pro	His Gly Cys Arg Leu Phe Tyr Ser Gln Leu Glu				
	260		265		270
Ala Thr Gln Glu	Gln Val Glu Leu Phe Gly Pro Ile Ser Leu Glu				
	275		280		285
Gln Val Arg Phe	Pro Ser Pro Glu Asp Ile Pro Ser Asp Lys Gln				
	290		295		300
Arg Phe Tyr Thr	Asn Gln Leu Leu Asp Val Leu Asp Arg Gly Leu				
	305		310		315
Ile Leu Gln Leu	Gln Gly Gln Asp Leu Tyr Ala Ile Arg Leu Cys				
	320		325		330
Gln Cys Lys Val	Phe Trp Ser Gly Pro Cys Ala Ser Ala His Asp				
	335		340		345
Ser Cys Pro Asn	Pro Ile Gln Arg Glu Val Lys Thr Lys Leu Phe				
	350		355		360
Ser Leu Glu His	Phe Leu Asn Glu Leu Ile Leu Phe Gln Lys Gly				
	365		370		375
Gln Thr Asn Thr	Pro Pro Pro Phe Glu Ile Phe Phe Cys Phe Gly				
	380		385		390
Glu Glu Trp Pro	Asp Arg Lys Pro Arg Glu Lys Lys Leu Ile Thr				
	395		400		405
Val Gln Val Val	Pro Val Ala Ala Arg Leu Leu Leu Glu Met Phe				
	410		415		420
Ser Gly Glu Leu	Ser Trp Ser Ala Asp Ser Ile Arg Leu Gln Ile				
	425		430		435
Ser Asn Pro Asp	Leu Lys Asp Arg Met Val Glu Gln Phe Lys Glu				
	440		445		450
Leu His His Ile	Trp Gln Ser Gln Gln Arg Leu Gln Pro Val Ala				
	455		460		465
Gln Ala Pro Pro	Gly Ala Gly Leu Gly Val Gly Gln Gly Pro Trp				
	470		475		480
Pro Met His Pro	Ala Gly Met Gln				
	485				

&lt;210&gt; 12

&lt;211&gt; 576

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7506086CD1

&lt;400&gt; 12

```

Met Leu Ser Leu Lys Lys Tyr Leu Thr Glu Gly Leu Leu Gln Phe
 1          5          10          15
Thr Ile Leu Leu Ser Leu Ile Gly Val Arg Val Asp Val Asp Thr
          20          25          30
Tyr Leu Thr Ser Gln Leu Pro Pro Leu Arg Glu Ile Ile Leu Gly
          35          40          45
Pro Ser Ser Ala Tyr Thr Gln Thr Gln Phe His Asn Leu Arg Asn
          50          55          60
Thr Leu Asp Gly Tyr Gly Ile His Pro Lys Ser Ile Asp Leu Asp
          65          70          75
Asn Tyr Phe Thr Ala Arg Arg Leu Leu Ser Gln Val Arg Ala Leu
          80          85          90
Asp Arg Phe Gln Val Pro Thr Thr Glu Val Asn Ala Trp Leu Val
          95          100          105
His Arg Asp Pro Glu Gly Ser Val Ser Gly Ser Gln Pro Asn Ser
          110          115          120
Gly Leu Ala Leu Glu Ser Ser Ser Gly Leu Gln Asp Val Thr Gly
          125          130          135
Pro Asp Asn Gly Val Arg Glu Ser Glu Thr Glu Gln Gly Phe Gly
          140          145          150
Glu Asp Leu Glu Asp Leu Gly Ala Val Ala Pro Pro Val Ser Gly
          155          160          165
Asp Leu Thr Lys Glu Asp Ile Asp Leu Ile Asp Ile Leu Trp Arg
          170          175          180
Gln Asp Ile Asp Leu Gly Ala Gly Arg Glu Val Phe Asp Tyr Ser
          185          190          195
His Arg Gln Lys Glu Gln Asp Val Glu Lys Glu Leu Arg Asp Gly
          200          205          210
Gly Glu Gln Asp Thr Trp Ala Gly Glu Gly Ala Glu Ala Leu Ala
          215          220          225
Arg Asn Leu Leu Val Asp Gly Glu Thr Gly Glu Ser Phe Pro Ala
          230          235          240
Gln Val Pro Ser Gly Glu Asp Gln Thr Ala Leu Ser Leu Glu Glu
          245          250          255
Cys Leu Arg Leu Leu Glu Ala Thr Cys Pro Phe Gly Glu Asn Ala
          260          265          270
Glu Phe Pro Ala Asp Ile Ser Ser Ile Thr Glu Ala Val Pro Ser
          275          280          285
Glu Ser Glu Pro Pro Ala Leu Gln Asn Asn Leu Leu Ser Pro Leu
          290          295          300
Leu Thr Gly Thr Glu Ser Pro Phe Asp Leu Glu Gln Gln Trp Gln
          305          310          315
Asp Leu Met Ser Ile Met Glu Met Gln Glu Gly Ala Val Gly Tyr
          320          325          330
Ser Ser Asp Ser Glu Thr Leu Asp Leu Glu Glu Ala Glu Gly Ala
          335          340          345
Val Gly Tyr Gln Pro Glu Tyr Ser Lys Phe Cys Arg Met Ser Tyr
          350          355          360
Gln Asp Pro Ala Gln Leu Ser Cys Leu Pro Tyr Leu Glu His Val
          365          370          375
Gly His Asn His Thr Tyr Asn Met Ala Pro Ser Ala Leu Asp Ser
          380          385          390

```

```

Ala Asp Leu Pro Pro Pro Ser Ala Leu Lys Lys Gly Ser Lys Glu
      395                      400                      405
Lys Gln Ala Asp Phe Leu Asp Lys Gln Met Ser Arg Asp Glu His
      410                      415                      420
Arg Ala Arg Ala Met Lys Ile Pro Phe Thr Asn Asp Lys Ile Ile
      425                      430                      435
Asn Leu Pro Val Glu Glu Phe Asn Glu Leu Leu Ser Lys Tyr Gln
      440                      445                      450
Leu Ser Glu Ala Gln Leu Ser Leu Ile Arg Asp Ile Arg Arg Arg
      455                      460                      465
Gly Lys Asn Lys Met Ala Ala Gln Asn Cys Arg Lys Arg Lys Leu
      470                      475                      480
Asp Thr Ile Leu Asn Leu Glu Arg Asp Val Glu Asp Leu Gln Arg
      485                      490                      495
Asp Lys Ala Arg Leu Leu Arg Glu Lys Val Glu Phe Leu Arg Ser
      500                      505                      510
Leu Arg Gln Met Lys Gln Lys Val Gln Ser Leu Tyr Gln Glu Val
      515                      520                      525
Phe Gly Arg Leu Arg Asp Glu Asn Gly Arg Pro Tyr Ser Pro Ser
      530                      535                      540
Gln Tyr Ala Leu Gln Tyr Ala Gly Asp Gly Ser Val Leu Leu Ile
      545                      550                      555
Pro Arg Thr Met Ala Asp Gln Gln Ala Arg Arg Gln Glu Arg Lys
      560                      565                      570
Pro Lys Asp Arg Arg Lys
      575

```

&lt;210&gt; 13

&lt;211&gt; 573

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 4176657CD1

&lt;400&gt; 13

```

Met Thr Pro Glu Ser Arg Asp Thr Thr Asp Leu Ser Pro Gly Gly
  1          5          10          15
Thr Gln Glu Met Glu Gly Ile Val Ile Val Lys Val Glu Glu Glu
  20          25          30
Asp Glu Glu Asp His Phe Gln Lys Glu Arg Asn Lys Val Glu Ser
  35          40          45
Ser Pro Gln Val Leu Ser Arg Ser Thr Thr Met Asn Glu Arg Ala
  50          55          60
Leu Leu Ser Ser Tyr Leu Val Ala Tyr Arg Val Ala Lys Glu Lys
  65          70          75
Met Ala His Thr Ala Ala Glu Lys Ile Ile Leu Pro Ala Cys Met
  80          85          90
Asp Met Val Arg Thr Ile Phe Asp Asp Lys Ser Ala Asp Lys Leu
  95          100         105
Arg Thr Ile Pro Leu Ser Asp Asn Thr Ile Ser Arg Arg Ile Cys
  110         115         120
Thr Ile Ala Lys His Leu Glu Ala Met Leu Ile Thr Arg Leu Gln
  125         130         135
Ser Gly Ile Asp Phe Ala Ile Gln Leu Asp Glu Ser Thr Asp Ile

```

	140		145		150
Ala Ser Cys Pro Thr	Leu Leu Val Tyr	Val Arg Tyr Val Trp	Gln		
	155		160		165
Asp Asp Phe Val Glu	Asp Leu Leu Cys	Cys Leu Asn Leu Asn	Ser		
	170		175		180
His Ile Thr Gly Leu	Asp Leu Phe Thr	Glu Leu Glu Asn Cys	Leu		
	185		190		195
Leu Gly Gln Tyr Lys	Leu Asn Trp Lys	His Cys Lys Gly Ile	Ser		
	200		205		210
Ser Asp Gly Thr Ala	Asn Met Thr Gly	Lys His Ser Arg Leu	Thr		
	215		220		225
Glu Lys Leu Leu Glu	Ala Thr His Asn	Asn Ala Val Trp Asn	His		
	230		235		240
Cys Phe Ile His Arg	Glu Ala Leu Val	Ser Lys Glu Ile Ser	Pro		
	245		250		255
Ser Leu Met Asp Val	Leu Lys Asn Ala	Val Lys Thr Val Asn	Phe		
	260		265		270
Ile Lys Gly Ser Ser	Leu Asn Ser Arg	Leu Leu Glu Ile Phe	Cys		
	275		280		285
Ser Glu Ile Gly Val	Asn His Thr His	Leu Leu Phe His Thr	Glu		
	290		295		300
Val Arg Trp Leu Ser	Gln Gly Lys Val	Leu Ser Arg Val Tyr	Glu		
	305		310		315
Leu Arg Asn Glu Ile	Tyr Ile Phe Leu	Val Glu Lys Gln Ser	His		
	320		325		330
Leu Ala Asn Ile Phe	Glu Asp Asp Ile	Trp Val Thr Lys Leu	Ala		
	335		340		345
Tyr Leu Ser Asp Ile	Phe Gly Ile Leu	Asn Glu Leu Ser Leu	Lys		
	350		355		360
Met Gln Gly Lys Asn	Asn Asp Ile Phe	Gln Tyr Leu Glu His	Ile		
	365		370		375
Leu Gly Phe Gln Lys	Thr Leu Leu Leu	Trp Gln Ala Arg Leu	Lys		
	380		385		390
Ser Asn Arg Pro Ser	Tyr Tyr Met Phe	Pro Thr Leu Leu Gln	His		
	395		400		405
Ile Glu Glu Asn Ile	Ile Asn Glu Asp	Cys Leu Lys Glu Ile	Lys		
	410		415		420
Leu Glu Ile Leu Leu	His Leu Thr Ser	Leu Ser Gln Thr Phe	Asn		
	425		430		435
Tyr Tyr Phe Pro Glu	Glu Lys Phe Glu	Ser Leu Lys Glu Asn	Ile		
	440		445		450
Trp Met Lys Asp Pro	Phe Ala Phe Gln	Asn Pro Glu Ser Ile	Ile		
	455		460		465
Glu Leu Asn Leu Glu	Pro Glu Glu Glu	Asn Glu Leu Leu Gln	Leu		
	470		475		480
Ser Ser Ser Phe Thr	Leu Lys Asn Tyr	Tyr Lys Ile Leu Ser	Leu		
	485		490		495
Ser Ala Phe Trp Ile	Lys Ile Lys Asp	Asp Phe Pro Leu Leu	Ser		
	500		505		510
Arg Lys Ser Ile Leu	Leu Leu Leu Leu	Pro Phe Thr Thr Thr	Tyr		
	515		520		525
Cys Glu Leu Gly Phe	Ser Ile Leu Thr	Arg Leu Lys Thr Lys	Lys		
	530		535		540
Arg Asn Arg Leu Asn	Ser Ala Pro Asp	Met Arg Val Ala Leu	Ser		
	545		550		555
Ser Cys Val Pro Asp	Trp Lys Glu Leu	Met Asn Arg Gln Ala	His		



Pro Ser His 560 565 570

<210> 14  
 <211> 501  
 <212> PRT  
 <213> Homo sapiens  
 <220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7506056CD1

<400> 14  
 Met Ala Pro Glu Arg Leu Arg Ser Arg Ala Leu Ser Ala Phe Lys  
 1 5 10 15  
 Leu Arg Gly Leu Leu Arg Gly Glu Ala Ile Lys Tyr Leu Thr  
 20 25 30  
 Glu Ala Leu Gln Ser Ile Ser Glu Leu Glu Leu Glu Asp Lys Leu  
 35 40 45  
 Glu Lys Ile Ile Asn Ala Val Glu Lys Gln Pro Leu Ser Ser Asn  
 50 55 60  
 Met Ile Glu Arg Ser Val Val Glu Ala Ala Val Gln Glu Cys Ser  
 65 70 75  
 Gln Ser Val Asp Glu Thr Ile Leu Leu Met Thr Asn His Pro Ala  
 80 85 90  
 Pro Asn Leu Phe Gly Thr Pro Arg Asp Lys Ala Glu Met Phe Arg  
 95 100 105  
 Glu Arg Tyr Thr Ile Leu His Gln Arg Thr His Arg His Glu Leu  
 110 115 120  
 Phe Thr Pro Pro Val Ile Gly Ser His Pro Asp Glu Ser Gly Ser  
 125 130 135  
 Lys Phe Gln Leu Lys Thr Ile Glu Thr Leu Leu Gly Ser Thr Thr  
 140 145 150  
 Lys Ile Gly Asp Ala Ile Val Leu Gly Met Ile Thr Gln Leu Lys  
 155 160 165  
 Glu Gly Lys Phe Phe Leu Glu Asp Pro Thr Gly Thr Val Gln Leu  
 170 175 180  
 Asp Leu Ser Lys Ala Gln Phe His Ser Gly Leu Tyr Thr Glu Ala  
 185 190 195  
 Cys Phe Val Leu Ala Glu Gly Trp Phe Glu Asp Gln Val Phe His  
 200 205 210  
 Val Asn Ala Phe Gly Phe Pro Pro Thr Glu Pro Ser Ser Thr Thr  
 215 220 225  
 Arg Ala Tyr Tyr Gly Asn Ile Asn Phe Phe Gly Gly Pro Ser Asn  
 230 235 240  
 Thr Ser Val Lys Thr Ser Ala Lys Leu Lys Gln Leu Glu Glu Glu  
 245 250 255  
 Asn Lys Asp Ala Met Phe Val Phe Leu Ser Asp Val Trp Leu Asp  
 260 265 270  
 Gln Val Glu Val Leu Glu Lys Leu Arg Ile Met Phe Ala Gly Tyr  
 275 280 285  
 Ser Pro Ala Pro Pro Thr Cys Phe Ile Leu Cys Gly Asn Phe Ser  
 290 295 300  
 Ser Ala Pro Tyr Gly Lys Asn Gln Val Gln Ala Leu Lys Asp Ser  
 305 310 315

```

Leu Lys Thr Leu Ala Asp Ile Ile Cys Glu Tyr Pro Asp Ile His
320 325 330
Gln Ser Ser Arg Phe Val Phe Val Pro Gly Pro Glu Asp Pro Gly
335 340 345
Phe Gly Ser Ile Leu Pro Arg Pro Pro Leu Ala Glu Ser Ile Thr
350 355 360
Asn Glu Phe Arg Gln Arg Val Pro Phe Ser Val Phe Thr Thr Asn
365 370 375
Pro Cys Arg Ile Gln Tyr Cys Thr Gln Glu Ile Thr Val Phe Arg
380 385 390
Glu Asp Leu Val Asn Lys Met Cys Arg Asn Cys Val Arg Phe Pro
395 400 405
Ser Ser Asn Leu Ala Ile Pro Asn His Phe Val Lys Thr Ile Leu
410 415 420
Ser Gln Gly His Leu Thr Pro Leu Pro Leu Tyr Val Cys Pro Val
425 430 435
Tyr Trp Ala Tyr Asp Tyr Ala Leu Arg Val Tyr Pro Val Pro Asp
440 445 450
Leu Leu Val Ile Ala Asp Lys Tyr Asp Pro Phe Thr Thr Thr Asn
455 460 465
Thr Glu Cys Leu Cys Ile Asn Pro Gly Ser Phe Pro Arg Ser Gly
470 475 480
Phe Ser Phe Lys Val Phe Tyr Pro Ser Asn Lys Thr Val Glu Asp
485 490 495
Ser Lys Leu Gln Gly Phe
500

```

&lt;210&gt; 15

&lt;211&gt; 893

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7506185CD1

&lt;400&gt; 15

```

Met Ala Asn Glu Thr Gln Lys Val Gly Ala Ile His Phe Pro Phe
1 5 10 15
Pro Phe Thr Pro Tyr Ser Ile Gln Glu Asp Phe Met Ala Glu Leu
20 25 30
Tyr Arg Val Leu Glu Ala Gly Lys Ile Gly Ile Phe Glu Ser Pro
35 40 45
Thr Gly Thr Gly Lys Ser Leu Ser Leu Ile Cys Gly Ala Leu Ser
50 55 60
Trp Leu Arg Asp Phe Glu Gln Lys Lys Arg Glu Glu Glu Ala Arg
65 70 75
Leu Leu Glu Thr Gly Thr Gly Pro Leu His Asp Glu Lys Asp Glu
80 85 90
Ser Leu Cys Leu Ser Ser Ser Cys Glu Gly Ala Ala Gly Thr Pro
95 100 105
Arg Pro Ala Gly Glu Pro Ala Trp Val Thr Gln Phe Val Gln Lys
110 115 120
Lys Glu Glu Arg Asp Leu Val Asp Arg Leu Lys Ala Glu Gln Ala
125 130 135
Arg Arg Lys Gln Arg Glu Glu Arg Leu Gln Gln Leu Gln His Arg

```

	140		145		150
Val Gln Leu Lys Tyr Ala Ala Lys Arg Leu Arg Gln Glu Glu Glu					
	155		160		165
Glu Arg Glu Asn Leu Leu Arg Leu Ser Arg Glu Met Leu Glu Thr					
	170		175		180
Gly Pro Glu Ala Glu Arg Leu Glu Gln Leu Glu Ser Gly Glu Glu					
	185		190		195
Glu Leu Val Leu Ala Glu Tyr Glu Ser Asp Glu Glu Lys Lys Val					
	200		205		210
Ala Ser Arg Val Asp Glu Asp Glu Asp Asp Leu Glu Glu Glu His					
	215		220		225
Ile Thr Lys Ile Tyr Tyr Cys Ser Arg Thr His Ser Gln Leu Ala					
	230		235		240
Gln Phe Val His Glu Val Lys Lys Ser Pro Phe Gly Lys Asp Val					
	245		250		255
Arg Leu Val Ser Leu Gly Ser Arg Gln Asn Leu Cys Val Asn Glu					
	260		265		270
Asp Val Lys Ser Leu Gly Ser Val Gln Leu Ile Asn Asp Arg Cys					
	275		280		285
Val Asp Met Gln Arg Ser Arg His Glu Lys Lys Lys Gly Ala Glu					
	290		295		300
Glu Glu Lys Pro Lys Arg Arg Arg Gln Glu Lys Gln Ala Ala Cys					
	305		310		315
Pro Phe Tyr Asn His Glu Gln Met Gly Leu Leu Arg Asp Glu Ala					
	320		325		330
Leu Ala Glu Val Lys Asp Met Glu Gln Leu Leu Ala Leu Gly Lys					
	335		340		345
Glu Ala Arg Ala Cys Pro Tyr Tyr Gly Ser Arg Leu Ala Ile Pro					
	350		355		360
Ala Ala Gln Leu Val Val Leu Pro Tyr Gln Met Leu Leu His Ala					
	365		370		375
Ala Thr Arg Gln Ala Ala Gly Ile Arg Leu Gln Asp Gln Val Val					
	380		385		390
Ile Ile Asp Glu Ala His Asn Leu Ile Asp Thr Ile Thr Gly Met					
	395		400		405
His Ser Val Glu Val Ser Gly Ser Gln Leu Cys Gln Ala His Ser					
	410		415		420
Gln Leu Leu Gln Tyr Val Glu Arg Tyr Gly Lys Arg Leu Lys Ala					
	425		430		435
Lys Asn Leu Met Tyr Leu Lys Gln Ile Leu Tyr Leu Leu Glu Lys					
	440		445		450
Phe Val Ala Val Leu Gly Gly Asn Ile Lys Gln Asn Pro Asn Thr					
	455		460		465
Gln Ser Leu Ser Gln Thr Gly Thr Glu Leu Lys Thr Ile Asn Asp					
	470		475		480
Phe Leu Phe Gln Ser Gln Ile Asp Asn Ile Asn Leu Phe Lys Val					
	485		490		495
Gln Arg Tyr Cys Glu Lys Ser Met Ile Ser Arg Lys Leu Phe Gly					
	500		505		510
Phe Thr Glu Arg Tyr Gly Ala Val Phe Ser Ser Arg Glu Gln Pro					
	515		520		525
Lys Leu Ala Gly Phe Gln Gln Phe Leu Gln Ser Leu Gln Pro Arg					
	530		535		540
Thr Thr Glu Ala Leu Ala Ala Pro Ala Asp Glu Ser Gln Ala Ser					
	545		550		555
Thr Leu Arg Pro Ala Ser Pro Leu Met His Ile Gln Gly Phe Leu					

560	565	570
Ala Ala Leu Thr Thr Ala Asn Gln Asp Gly Arg Val Ile Leu Ser		
575	580	585
Arg Gln Gly Ser Leu Ser Gln Ser Thr Leu Lys Phe Leu Leu Leu		
590	595	600
Asn Pro Ala Val His Phe Ala Gln Val Val Lys Glu Cys Arg Ala		
605	610	615
Val Val Ile Ala Gly Gly Thr Met Gln Pro Val Ser Asp Phe Arg		
620	625	630
Gln Gln Leu Leu Ala Cys Ala Gly Val Glu Ala Glu Arg Val Val		
635	640	645
Glu Phe Ser Cys Gly His Val Ile Pro Pro Asp Asn Ile Leu Pro		
650	655	660
Leu Val Ile Cys Ser Gly Ile Ser Asn Gln Pro Leu Glu Phe Thr		
665	670	675
Phe Gln Lys Arg Glu Leu Pro Gln Met Met Asp Glu Val Gly Arg		
680	685	690
Ile Leu Cys Asn Leu Cys Gly Val Val Pro Gly Gly Val Val Cys		
695	700	705
Phe Phe Pro Ser Tyr Glu Tyr Leu Arg Gln Val His Ala His Trp		
710	715	720
Glu Lys Gly Gly Leu Leu Gly Arg Leu Ala Ala Arg Lys Lys Ile		
725	730	735
Phe Gln Glu Pro Lys Ser Ala His Gln Val Glu Gln Val Leu Leu		
740	745	750
Ala Tyr Ser Arg Cys Ile Gln Ala Cys Gly Gln Glu Arg Gly Gln		
755	760	765
Val Thr Gly Ala Leu Leu Leu Ser Val Val Gly Gly Lys Met Ser		
770	775	780
Glu Gly Ile Asn Phe Ser Asp Asn Leu Gly Arg Ser Ala Glu Leu		
785	790	795
Gln Glu Lys Met Ala Tyr Leu Asp Gln Thr Leu Pro Arg Ala Pro		
800	805	810
Gly Gln Ala Pro Pro Gly Lys Ala Leu Val Glu Asn Leu Cys Met		
815	820	825
Lys Ala Val Asn Gln Ser Ile Gly Arg Ala Ile Arg His Gln Lys		
830	835	840
Asp Phe Ala Ser Val Val Leu Leu Asp Gln Arg Tyr Ala Arg Pro		
845	850	855
Pro Val Leu Ala Lys Leu Pro Ala Trp Ile Arg Ala Arg Val Glu		
860	865	870
Val Lys Ala Thr Phe Gly Pro Ala Ile Ala Ala Val Gln Lys Phe		
875	880	885
His Arg Glu Lys Ser Ala Ser Ser		
890		

&lt;210&gt; 16

&lt;211&gt; 555

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 8096611CD1

&lt;400&gt; 16

Met	Phe	Tyr	Thr	His	Val	Leu	Met	Ser	Lys	Arg	Gly	Pro	Leu	Ala
1				5					10					15
Lys	Ile	Trp	Leu	Ala	Ala	His	Trp	Glu	Lys	Lys	Leu	Thr	Lys	Ala
				20					25					30
His	Val	Phe	Glu	Cys	Asn	Leu	Glu	Ile	Thr	Ile	Glu	Lys	Ile	Leu
				35					40					45
Ser	Pro	Lys	Val	Lys	Ile	Ala	Leu	Arg	Thr	Ser	Gly	His	Leu	Leu
				50					55					60
Leu	Gly	Val	Val	Arg	Ile	Tyr	Asn	Arg	Lys	Ala	Lys	Tyr	Leu	Leu
				65					70					75
Ala	Asp	Cys	Ser	Glu	Ala	Phe	Leu	Lys	Met	Lys	Met	Thr	Phe	Arg
				80					85					90
Pro	Gly	Leu	Val	Asp	Leu	Pro	Lys	Glu	Asn	Phe	Glu	Ala	Ser	Tyr
				95					100					105
Asn	Ala	Ile	Thr	Leu	Pro	Glu	Glu	Phe	His	Asp	Phe	Asp	Thr	Gln
				110					115					120
Asn	Met	Asn	Ala	Ile	Asp	Val	Ser	Glu	His	Phe	Thr	Gln	Asn	Gln
				125					130					135
Ser	Arg	Pro	Glu	Glu	Ile	Thr	Leu	Arg	Glu	Asn	Leu	Asp	Asn	Asp
				140					145					150
Leu	Leu	Phe	Gln	Ala	Glu	Ser	Phe	Gly	Glu	Glu	Ser	Glu	Ile	Leu
				155					160					165
Arg	Arg	His	Ser	Phe	Phe	Asp	Asp	Asn	Ile	Leu	Leu	Asn	Ser	Ser
				170					175					180
Gly	Pro	Leu	Ile	Glu	His	Ser	Ser	Gly	Ser	Leu	Thr	Gly	Glu	Arg
				185					190					195
Ser	Leu	Phe	Tyr	Asp	Ser	Gly	Asp	Gly	Phe	Gly	Asp	Glu	Gly	Ala
				200					205					210
Ala	Gly	Glu	Met	Ile	Asp	Asn	Leu	Leu	Gln	Asp	Asp	Gln	Asn	Ile
				215					220					225
Leu	Leu	Glu	Asp	Met	His	Leu	Asn	Arg	Glu	Ile	Ser	Leu	Pro	Ser
				230					235					240
Glu	Pro	Pro	Asn	Ser	Leu	Ala	Val	Glu	Pro	Asp	Asn	Ser	Glu	Cys
				245					250					255
Ile	Cys	Val	Pro	Glu	Asn	Glu	Lys	Met	Asn	Glu	Thr	Ile	Leu	Leu
				260					265					270
Ser	Thr	Glu	Glu	Glu	Gly	Phe	Thr	Leu	Asp	Pro	Ile	Asp	Ile	Ser
				275					280					285
Asp	Ile	Ala	Glu	Lys	Arg	Lys	Gly	Lys	Lys	Arg	Arg	Leu	Leu	Ile
				290					295					300
Asp	Pro	Ile	Lys	Glu	Leu	Ser	Ser	Lys	Val	Ile	His	Lys	Gln	Leu
				305					310					315
Thr	Ser	Phe	Ala	Asp	Thr	Leu	Met	Val	Leu	Glu	Leu	Ala	Pro	Pro
				320					325					330
Thr	Gln	Arg	Leu	Met	Met	Trp	Lys	Lys	Arg	Gly	Gly	Val	His	Thr
				335					340					345
Leu	Leu	Ser	Thr	Ala	Ala	Gln	Asp	Leu	Ile	His	Ala	Glu	Leu	Lys
				350					355					360
Met	Leu	Phe	Thr	Lys	Cys	Phe	Leu	Ser	Ser	Gly	Phe	Lys	Leu	Gly
				365					370					375
Arg	Lys	Met	Ile	Gln	Lys	Glu	Ser	Val	Arg	Glu	Glu	Val	Gly	Asn
				380					385					390
Gln	Asn	Ile	Val	Glu	Thr	Ser	Met	Met	Gln	Glu	Pro	Asn	Tyr	Gln
				395					400					405
Gln	Glu	Leu	Ser	Lys	Pro	Gln	Thr	Trp	Lys	Asp	Val	Ile	Gly	Gly
				410					415					420

Ser	Gln	His	Ser	Ser	His	Glu	Asp	Thr	Asn	Lys	Asn	Ile	Asn	Ser
				425					430					435
Glu	Asp	Ile	Val	Glu	Met	Val	Ser	Leu	Ala	Ala	Glu	Glu	Ser	Ser
				440					445					450
Leu	Met	Asn	Asp	Leu	Phe	Ala	Gln	Glu	Ile	Glu	Tyr	Ser	Pro	Val
				455					460					465
Glu	Leu	Glu	Ser	Leu	Ser	Asn	Glu	Glu	Asn	Ile	Glu	Thr	Glu	Arg
				470					475					480
Trp	Asn	Gly	Arg	Ile	Leu	Gln	Met	Leu	Asn	Arg	Leu	Arg	Glu	Ser
				485					490					495
Asn	Lys	Met	Gly	Met	Gln	Ser	Phe	Ser	Leu	Met	Lys	Leu	Cys	Arg
				500					505					510
Asn	Ser	Asp	Arg	Lys	Gln	Ala	Ala	Ala	Lys	Phe	Tyr	Ser	Phe	Leu
				515					520					525
Val	Leu	Lys	Lys	Gln	Leu	Ala	Ile	Glu	Leu	Ser	Gln	Ser	Ala	Pro
				530					535					540
Tyr	Ala	Asp	Ile	Ile	Ala	Thr	Met	Gly	Pro	Met	Phe	Tyr	Asn	Ile
				545					550					555

&lt;210&gt; 17

&lt;211&gt; 584

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 8174603CD1

&lt;400&gt; 17

Met	Ala	Ser	Ala	Ser	Gln	Gly	Ala	Asp	Asp	Asp	Gly	Ser	Arg	Arg
1				5					10					15
Lys	Pro	Arg	Leu	Ala	Ala	Ser	Leu	Gln	Ile	Ser	Pro	Gln	Pro	Arg
				20					25					30
Pro	Trp	Arg	Pro	Leu	Pro	Ala	Gln	Ala	Gln	Ser	Ala	Trp	Glu	Pro
				35					40					45
Cys	Ala	Arg	Cys	Pro	Arg	Arg	Ile	Ala	Ala	Glu	Gly	Gly	Trp	Pro
				50					55					60
Gln	Val	Ser	Val	Leu	Arg	Asp	Ser	Gly	Pro	Gly	Ala	Gly	Ala	Gly
				65					70					75
Val	Gly	Glu	Leu	Gly	Ala	Ala	Arg	Ala	Trp	Glu	Asn	Leu	Gly	Glu
				80					85					90
Gln	Met	Gly	Lys	Ala	Pro	Arg	Val	Pro	Val	Pro	Pro	Ala	Gly	Leu
				95					100					105
Ser	Leu	Pro	Leu	Lys	Asp	Pro	Pro	Ala	Ser	Gln	Ala	Val	Ser	Leu
				110					115					120
Leu	Thr	Glu	Tyr	Ala	Ala	Ser	Leu	Gly	Ile	Phe	Leu	Leu	Phe	Arg
				125					130					135
Glu	Asp	Gln	Pro	Pro	Gly	Glu	Ala	Gly	Pro	Cys	Phe	Pro	Phe	Ser
				140					145					150
Val	Ser	Ala	Glu	Leu	Asp	Gly	Val	Val	Cys	Pro	Ala	Gly	Thr	Ala
				155					160					165
Asn	Ser	Lys	Thr	Glu	Ala	Lys	Gln	Gln	Ala	Thr	Leu	Ser	Ala	Leu
				170					175					180
Cys	Tyr	Ile	Arg	Ser	Gln	Leu	Glu	Asn	Pro	Val	Gly	Pro	Leu	Leu
				185					190					195

```

Pro Ala Val Ser Arg Pro Gly Ala Glu Asn Ile Leu Thr His Glu
200 205 210
Gln Arg Cys Ala Ala Leu Val Ser Ala Gly Phe Asp Leu Leu Leu
215 220 225
Asp Glu Arg Ser Pro Tyr Trp Ala Cys Lys Gly Thr Val Ala Gly
230 235 240
Val Ile Leu Glu Arg Glu Ile Pro Arg Ala Arg Gly His Val Lys
245 250 255
Glu Ile Tyr Lys Leu Val Ala Leu Gly Thr Gly Ser Ser Cys Cys
260 265 270
Ala Gly Trp Leu Glu Phe Ser Gly Gln Gln Leu His Asp Cys His
275 280 285
Gly Leu Val Ile Ala Arg Arg Ala Leu Leu Arg Phe Leu Phe Arg
290 295 300
Gln Leu Leu Leu Ala Thr Gln Gly Gly Pro Lys Gly Lys Glu Gln
305 310 315
Ser Val Leu Ala Pro Gln Pro Gly Pro Gly Pro Pro Phe Thr Leu
320 325 330
Lys Pro Arg Val Phe Leu His Leu Tyr Ile Ser Asn Thr Pro Lys
335 340 345
Gly Ala Ala Arg Asp Ile Tyr Leu Pro Pro Thr Ser Glu Gly Gly
350 355 360
Leu Pro His Ser Pro Pro Met Arg Leu Gln Ala His Val Leu Gly
365 370 375
Gln Leu Lys Pro Val Cys Tyr Val Ala Pro Ser Leu Cys Asp Thr
380 385 390
His Val Gly Cys Leu Ser Ala Ser Asp Lys Leu Ala Arg Trp Ala
395 400 405
Val Leu Gly Leu Gly Gly Ala Leu Leu Ala His Leu Val Ser Pro
410 415 420
Leu Tyr Ser Thr Ser Leu Ile Leu Ala Asp Ser Cys His Asp Pro
425 430 435
Pro Thr Leu Ser Arg Ala Ile His Thr Arg Pro Cys Leu Asp Ser
440 445 450
Val Leu Gly Pro Cys Leu Pro Pro Pro Tyr Val Arg Thr Ala Leu
455 460 465
His Leu Phe Ala Gly Pro Pro Val Ala Pro Ser Glu Pro Thr Pro
470 475 480
Asp Thr Cys Arg Gly Leu Ser Leu Asn Trp Ser Leu Gly Asp Pro
485 490 495
Gly Ile Glu Val Val Asp Val Ala Thr Gly Arg Val Lys Ala Asn
500 505 510
Ala Ala Leu Gly Pro Pro Ser Arg Leu Cys Lys Ala Ser Phe Leu
515 520 525
Arg Ala Phe His Gln Ala Ala Arg Ala Val Gly Lys Pro Tyr Leu
530 535 540
Leu Ala Leu Lys Thr Tyr Glu Ala Ala Lys Ala Gly Pro Tyr Gln
545 550 555
Glu Ala Arg Arg Gln Leu Ser Leu Leu Leu Asp Gln Gln Gly Leu
560 565 570
Gly Ala Trp Pro Ser Lys Pro Leu Val Gly Lys Phe Arg Asn
575 580

```

<210> 18  
 <211> 554  
 <212> PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 3101042CD1

&lt;400&gt; 18

```

Met Leu Glu Asn Phe Arg Asn Leu Met Leu Val Arg Asp Gly Ile
 1           5           10           15
Lys Asn Asn Ile Leu Asn Leu Gln Ala Lys Gly Leu Ser Tyr Leu
      20           25           30
Ser Gln Glu Val Leu His Cys Trp Gln Ile Trp Lys Gln Arg Ile
      35           40           45
Arg Asp Leu Thr Val Ser Gln Asp Tyr Ile Val Asn Leu Gln Glu
      50           55           60
Glu Cys Ser Pro His Leu Glu Asp Val Ser Leu Ser Glu Glu Trp
      65           70           75
Ala Gly Ile Ser Leu Gln Ile Ser Glu Asn Glu Asn Tyr Val Val
      80           85           90
Asn Ala Ile Ile Lys Asn Gln Asp Ile Thr Ala Trp Gln Ser Leu
      95          100          105
Thr Gln Val Leu Thr Pro Glu Ser Trp Arg Lys Ala Asn Ile Met
     110          115          120
Thr Glu Pro Gln Asn Ser Gln Gly Arg Tyr Lys Gly Ile Tyr Met
     125          130          135
Glu Glu Lys Leu Tyr Arg Arg Ala Gln His Asp Asp Ser Leu Ser
     140          145          150
Trp Thr Ser Cys Asp His His Glu Ser Gln Glu Cys Lys Gly Glu
     155          160          165
Asp Pro Gly Arg His Pro Ser Cys Gly Lys Asn Leu Gly Met Lys
     170          175          180
Ser Thr Val Glu Lys Arg Asn Ala Ala His Val Leu Pro Gln Pro
     185          190          195
Phe Pro Cys Asn Asn Cys Gly Val Ala Phe Ala Asp Asp Thr Asp
     200          205          210
Pro His Val His His Ser Thr His Leu Gly Glu Lys Ser Tyr Lys
     215          220          225
Cys Asp Gln Tyr Gly Lys Asn Phe Ser Gln Ser Gln Asp Leu Ile
     230          235          240
Val His Cys Lys Thr His Ser Gly Lys Thr Pro Tyr Glu Phe His
     245          250          255
Glu Trp Pro Met Gly Cys Lys Gln Ser Ser Asp Leu Pro Arg Tyr
     260          265          270
Gln Lys Val Ser Ser Gly Asp Lys Pro Tyr Lys Cys Lys Glu Cys
     275          280          285
Gly Lys Gly Phe Arg Arg Ser Ser Ser Leu His Asn His His Arg
     290          295          300
Val His Thr Gly Glu Met Pro Tyr Lys Cys Asp Glu Cys Gly Lys
     305          310          315
Gly Phe Gly Phe Arg Ser Leu Leu Cys Ile His Gln Gly Val His
     320          325          330
Thr Gly Lys Lys Pro Tyr Lys Cys Glu Glu Cys Gly Lys Gly Phe
     335          340          345
Asp Gln Ser Ser Asn Leu Leu Val His Gln Arg Val His Thr Gly
     350          355          360
Glu Lys Pro Tyr Lys Cys Ser Glu Cys Gly Lys Cys Phe Ser Ser

```



	365		370		375
Ser Ser Val Leu	Gln Val His Trp Arg	Phe His Thr Gly Glu Lys			
	380		385		390
Pro Tyr Arg Cys	Gly Glu Cys Gly Lys	Gly Phe Ser Gln Cys Thr			
	395		400		405
His Leu His Ile	His Gln Arg Val His	Thr Gly Glu Lys Pro Tyr			
	410		415		420
Lys Cys Asn Val	Cys Gly Lys Asp Phe	Ala Tyr Ser Ser Val Leu			
	425		430		435
His Thr His Gln	Arg Val His Thr Gly	Glu Lys Pro Tyr Lys Cys			
	440		445		450
Glu Val Cys Gly	Lys Cys Phe Ser Tyr	Ser Ser Tyr Phe His Leu			
	455		460		465
His Gln Arg Asp	His Ile Arg Glu Lys	Pro Tyr Lys Cys Asp Glu			
	470		475		480
Cys Gly Lys Gly	Phe Ser Arg Asn Ser	Asp Leu Asn Val His Leu			
	485		490		495
Arg Val His Thr	Arg Glu Arg Pro Tyr	Lys Cys Lys Ala Cys Gly			
	500		505		510
Lys Gly Phe Ser	Arg Asn Ser Tyr Leu	Leu Ala His Gln Arg Val			
	515		520		525
His Ile Asp Glu	Thr Gln Tyr Thr His	Cys Glu Arg Gly Lys Asp			
	530		535		540
Leu Leu Thr His	Gln Arg Leu His Glu	Gln Arg Glu Thr Leu			
	545		550		

&lt;210&gt; 19

&lt;211&gt; 1004

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 4972035CD1

&lt;400&gt; 19

Met Trp Glu Lys Met	Glu Thr Lys Thr	Ile Val Tyr Asp Leu Asp
1	5	10 15
Thr Ser Gly Gly	Leu Met Glu Gln Ile	Gln Ala Leu Leu Ala Pro
	20	25 30
Pro Lys Thr Asp	Glu Ala Glu Lys Arg	Ser Arg Lys Pro Glu Lys
	35	40 45
Glu Pro Arg Arg	Ser Gly Arg Ala Thr	Asn His Asp Ser Cys Asp
	50	55 60
Ser Cys Lys Glu	Gly Gly Asp Leu Leu	Cys Cys Asp His Cys Pro
	65	70 75
Ala Ala Phe His	Leu Gln Cys Cys Asn	Pro Pro Leu Ser Glu Glu
	80	85 90
Met Leu Pro Pro	Gly Glu Trp Met Cys	His Arg Cys Thr Val Arg
	95	100 105
Arg Lys Lys Arg	Gln Lys Lys Glu	Leu Gly His Val Asn Gly
	110	115 120
Leu Val Asp Lys	Ser Gly Lys Arg Thr	Thr Ser Pro Ser Ser Asp
	125	130 135
Thr Asp Leu Leu	Asp Arg Ser Ala Ser	Lys Thr Glu Leu Lys Ala
	140	145 150

Ile	Ala	His	Ala	Arg	Ile	Leu	Glu	Arg	Arg	Ala	Ser	Arg	Pro	Gly
				155					160					165
Thr	Pro	Thr	Ser	Ser	Ala	Ser	Thr	Glu	Thr	Pro	Thr	Ser	Glu	Gln
				170					175					180
Asn	Asp	Val	Asp	Glu	Asp	Ile	Ile	Asp	Val	Asp	Glu	Glu	Pro	Val
				185					190					195
Ala	Ala	Glu	Pro	Asp	Tyr	Val	Gln	Pro	Gln	Leu	Arg	Arg	Pro	Phe
				200					205					210
Glu	Leu	Leu	Ile	Ala	Ala	Ala	Met	Glu	Arg	Asn	Pro	Thr	Gln	Phe
				215					220					225
Gln	Leu	Pro	Asn	Glu	Leu	Thr	Cys	Thr	Thr	Ala	Leu	Pro	Gly	Ser
				230					235					240
Ser	Lys	Arg	Arg	Arg	Lys	Glu	Glu	Thr	Thr	Gly	Lys	Asn	Val	Lys
				245					250					255
Lys	Thr	Gln	His	Glu	Leu	Asp	His	Asn	Gly	Leu	Val	Pro	Leu	Pro
				260					265					270
Val	Lys	Val	Cys	Phe	Thr	Cys	Asn	Arg	Ser	Cys	Arg	Val	Ala	Pro
				275					280					285
Leu	Ile	Gln	Cys	Asp	Tyr	Cys	Pro	Leu	Leu	Phe	His	Met	Asp	Cys
				290					295					300
Leu	Glu	Pro	Pro	Leu	Thr	Ala	Met	Pro	Leu	Gly	Arg	Trp	Met	Cys
				305					310					315
Pro	Asn	His	Ile	Glu	His	Val	Val	Leu	Asn	Gln	Lys	Asn	Met	Thr
				320					325					330
Leu	Ser	Asn	Arg	Cys	Gln	Val	Phe	Asp	Arg	Phe	Gln	Asp	Thr	Val
				335					340					345
Ser	Gln	His	Val	Val	Lys	Val	Asp	Phe	Leu	Asn	Arg	Ile	His	Lys
				350					355					360
Lys	His	Pro	Pro	Asn	Arg	Arg	Val	Leu	Gln	Ser	Val	Lys	Arg	Arg
				365					370					375
Ser	Leu	Lys	Val	Pro	Asp	Ala	Ile	Lys	Ser	Gln	Tyr	Gln	Phe	Pro
				380					385					390
Pro	Pro	Leu	Ile	Ala	Pro	Ala	Ala	Ile	Arg	Asp	Gly	Glu	Leu	Ile
				395					400					405
Cys	Asn	Gly	Ile	Pro	Glu	Glu	Ser	Gln	Met	His	Leu	Leu	Asn	Ser
				410					415					420
Glu	His	Leu	Ala	Thr	Gln	Ala	Glu	Gln	Gln	Glu	Trp	Leu	Cys	Ser
				425					430					435
Val	Val	Ala	Leu	Gln	Cys	Ser	Ile	Leu	Lys	His	Leu	Ser	Ala	Lys
				440					445					450
Gln	Met	Pro	Ser	His	Trp	Asp	Ser	Glu	Gln	Thr	Glu	Lys	Ala	Asp
				455					460					465
Ile	Lys	Pro	Val	Ile	Val	Thr	Asp	Ser	Ser	Val	Thr	Thr	Ser	Leu
				470					475					480
Gln	Thr	Ala	Asp	Lys	Thr	Pro	Thr	Pro	Ser	His	Tyr	Pro	Leu	Ser
				485					490					495
Cys	Pro	Ser	Gly	Ile	Ser	Thr	Gln	Asn	Ser	Leu	Ser	Cys	Ser	Pro
				500					505					510
Pro	His	Gln	Ser	Pro	Ala	Leu	Glu	Asp	Ile	Gly	Cys	Ser	Ser	Cys
				515					520					525
Ala	Glu	Lys	Ser	Lys	Thr	Pro	Cys	Gly	Thr	Ala	Asn	Gly	Pro	
				530					535					540
Val	Asn	Thr	Glu	Val	Lys	Ala	Asn	Gly	Pro	His	Leu	Tyr	Ser	Ser
				545					550					555
Pro	Thr	Asp	Ser	Thr	Asp	Pro	Arg	Arg	Leu	Pro	Gly	Ala	Asn	Thr
				560					565					570

Pro	Leu	Pro	Gly	Leu	Ser	His	Arg	Gln	Gly	Trp	Pro	Arg	Pro	Leu	575	580	585
Thr	Pro	Pro	Ala	Ala	Gly	Gly	Leu	Gln	Asn	His	Thr	Val	Gly	Ile	590	595	600
Ile	Val	Lys	Thr	Glu	Asn	Ala	Thr	Gly	Pro	Ser	Ser	Cys	Pro	Gln	605	610	615
Arg	Ser	Leu	Val	Pro	Val	Pro	Ser	Leu	Pro	Pro	Ser	Ile	Pro	Ser	620	625	630
Ser	Cys	Ala	Ser	Ile	Glu	Asn	Thr	Ser	Thr	Leu	Gln	Arg	Lys	Thr	635	640	645
Val	Gln	Ser	Gln	Ile	Gly	Pro	Pro	Leu	Thr	Asp	Ser	Arg	Pro	Leu	650	655	660
Gly	Ser	Pro	Pro	Asn	Ala	Thr	Arg	Val	Leu	Thr	Pro	Pro	Gln	Ala	665	670	675
Ala	Gly	Asp	Gly	Ile	Leu	Ala	Thr	Thr	Ala	Asn	Gln	Arg	Phe	Ser	680	685	690
Ser	Pro	Ala	Pro	Ser	Ser	Asp	Gly	Lys	Val	Ser	Pro	Gly	Thr	Leu	695	700	705
Ser	Ile	Gly	Ser	Ala	Leu	Thr	Val	Pro	Ser	Phe	Pro	Ala	Asn	Ser	710	715	720
Thr	Ala	Met	Val	Asp	Leu	Thr	Asn	Ser	Leu	Arg	Ala	Phe	Met	Asp	725	730	735
Val	Asn	Gly	Glu	Ile	Glu	Ile	Asn	Met	Leu	Asp	Glu	Lys	Leu	Ile	740	745	750
Lys	Phe	Leu	Ala	Leu	Gln	Arg	Ile	His	Gln	Leu	Phe	Pro	Ser	Arg	755	760	765
Val	Gln	Pro	Ser	Pro	Gly	Ser	Val	Gly	Thr	His	Gln	Leu	Ala	Ser	770	775	780
Gly	Gly	His	His	Ile	Glu	Val	Gln	Arg	Lys	Glu	Val	Gln	Ala	Arg	785	790	795
Ala	Val	Phe	Tyr	Pro	Leu	Leu	Gly	Leu	Gly	Gly	Ala	Val	Asn	Met	800	805	810
Cys	Tyr	Arg	Thr	Leu	Tyr	Ile	Gly	Thr	Gly	Ala	Asp	Met	Asp	Val	815	820	825
Cys	Leu	Thr	Asn	Tyr	Gly	His	Cys	Asn	Tyr	Val	Ser	Gly	Lys	His	830	835	840
Ala	Cys	Ile	Phe	Tyr	Asp	Glu	Asn	Thr	Lys	His	Tyr	Glu	Leu	Leu	845	850	855
Asn	Tyr	Ser	Glu	His	Gly	Thr	Thr	Val	Asp	Asn	Val	Leu	Tyr	Ser	860	865	870
Cys	Asp	Phe	Ser	Glu	Lys	Thr	Pro	Pro	Thr	Pro	Pro	Ser	Ser	Ile	875	880	885
Val	Ala	Lys	Val	Gln	Ser	Val	Ile	Arg	Arg	Arg	Arg	His	Gln	Lys	890	895	900
Gln	Asp	Glu	Glu	Pro	Ser	Glu	Glu	Ala	Ala	Met	Met	Ser	Ser	Gln	905	910	915
Ala	Gln	Gly	Pro	Gln	Arg	Arg	Pro	Cys	Asn	Cys	Lys	Ala	Ser	Ser	920	925	930
Ser	Ser	Leu	Ile	Gly	Gly	Ser	Gly	Ala	Gly	Trp	Glu	Gly	Thr	Ala	935	940	945
Leu	Leu	His	His	Gly	Ser	Tyr	Ile	Lys	Leu	Gly	Cys	Leu	Gln	Phe	950	955	960
Val	Phe	Ser	Ile	Thr	Glu	Phe	Ala	Thr	Lys	Gln	Pro	Lys	Gly	Asp	965	970	975
Ala	Ser	Leu	Leu	Gln	Asp	Gly	Val	Leu	Ala	Glu	Lys	Leu	Ser	Leu	980	985	990

Lys Pro His Gln Gly Pro Val Leu Arg Ser Asn Ser Val Pro  
 995 1000

<210> 20  
 <211> 123  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7506265CD1

<400> 20  
 Met Ser Leu Val Ile Pro Glu Lys Phe Gln His Ile Leu Arg Val  
 1 5 10 15  
 Leu Asn Thr Asn Ile Asp Gly Arg Arg Lys Ile Ala Phe Ala Ile  
 20 25 30  
 Thr Ala Ile Lys Val Glu Arg Val Ile Thr Ile Met Gln Asn Pro  
 35 40 45  
 Arg Gln Tyr Lys Ile Pro Asp Trp Phe Leu Asn Arg Gln Lys Asp  
 50 55 60  
 Val Lys Asp Gly Lys Tyr Ser Gln Val Leu Ala Asn Gly Leu Asp  
 65 70 75  
 Asn Lys Leu Arg Glu Asp Leu Glu Arg Leu Lys Lys Ile Arg Ala  
 80 85 90  
 His Arg Gly Leu Arg His Phe Trp Gly Leu Arg Val Arg Gly Gln  
 95 100 105  
 His Thr Lys Thr Gly Arg Arg Gly Arg Thr Val Gly Val Ser  
 110 115 120  
 Lys Lys Lys

<210> 21  
 <211> 112  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7506304CD1

<400> 21  
 Met Ser Ala His Leu Gln Trp Met Val Val Arg Asn Cys Ser Ser  
 1 5 10 15  
 Phe Leu Ile Lys Arg Asn Lys Gln Thr Tyr Ser Thr Glu Pro Asn  
 20 25 30  
 Asn Leu Lys Ala Arg Asn Ser Phe Arg Tyr Asn Gly Leu Ile His  
 35 40 45  
 Arg Lys Thr Ala Thr Ser Tyr Val Arg Thr Thr Ile Asn Lys Asn  
 50 55 60  
 Ala Arg Ala Thr Leu Ser Ser Ile Arg His Met Ile Arg Lys Asn  
 65 70 75  
 Lys Tyr Arg Pro Asp Leu Arg Met Ala Ala Ile Arg Arg Ala Ser  
 80 85 90  
 Ala Ile Leu Arg Ser Gln Lys Pro Val Met Val Lys Arg Lys Arg  
 95 100 105

Thr Arg Pro Thr Lys Ser Ser  
110

<210> 22

<211> 987

<212> PRT

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 7506198CD1

<400> 22

Met	Asn	His	Asp	Phe	Gln	Ala	Leu	Ala	Leu	Glu	Ser	Arg	Gly	Met
1				5					10					15
Gly	Glu	Leu	Leu	Pro	Thr	Lys	Lys	Phe	Trp	Glu	Pro	Asp	Asp	Ser
				20					25					30
Thr	Lys	Asp	Gly	Gln	Lys	Gly	Ile	Phe	Leu	Gly	Asp	Asp	Glu	Trp
				35					40					45
Arg	Glu	Thr	Ala	Trp	Gly	Ala	Ser	His	His	Ser	Met	Ser	Gln	Pro
				50					55					60
Ile	Met	Val	Gln	Arg	Arg	Ser	Gly	Gln	Gly	Phe	His	Gly	Asn	Ser
				65					70					75
Glu	Val	Asn	Ala	Ile	Leu	Ser	Pro	Arg	Ser	Glu	Ser	Gly	Gly	Leu
				80					85					90
Gly	Val	Ser	Met	Val	Glu	Tyr	Val	Leu	Ser	Ser	Ser	Pro	Ala	Asp
				95					100					105
Lys	Leu	Asp	Ser	Arg	Phe	Arg	Lys	Gly	Asn	Phe	Gly	Thr	Arg	Asp
				110					115					120
Ala	Glu	Thr	Asp	Gly	Pro	Glu	Lys	Gly	Asp	Gln	Lys	Gly	Lys	Ala
				125					130					135
Ser	Pro	Phe	Glu	Glu	Asp	Gln	Asn	Arg	Asp	Leu	Lys	Gln	Gly	Asp
				140					145					150
Asp	Asp	Asp	Ser	Lys	Ile	Asn	Gly	Arg	Gly	Leu	Pro	Asn	Gly	Met
				155					160					165
Asp	Ala	Asp	Cys	Lys	Asp	Phe	Asn	Arg	Thr	Pro	Gly	Ser	Arg	Gln
				170					175					180
Ala	Ser	Pro	Thr	Glu	Val	Val	Glu	Arg	Leu	Gly	Pro	Asn	Thr	Asn
				185					190					195
Pro	Ser	Glu	Gly	Leu	Gly	Pro	Leu	Pro	Asn	Pro	Thr	Ala	Asn	Lys
				200					205					210
Pro	Leu	Val	Glu	Glu	Phe	Ser	Asn	Pro	Glu	Thr	Gln	Asn	Leu	Asp
				215					220					225
Ala	Met	Glu	Gln	Val	Gly	Leu	Glu	Ser	Leu	Gln	Phe	Asp	Tyr	Pro
				230					235					240
Gly	Asn	Gln	Val	Pro	Met	Asp	Ser	Ser	Gly	Ala	Thr	Val	Gly	Leu
				245					250					255
Phe	Asp	Tyr	Asn	Ser	Gln	Gln	Gln	Leu	Phe	Gln	Arg	Thr	Asn	Ala
				260					265					270
Leu	Thr	Val	Gln	Gln	Leu	Thr	Ala	Ala	Gln	Gln	Gln	Gln	Tyr	Ala
				275					280					285
Leu	Ala	Ala	Ala	Gln	Gln	Pro	His	Ile	Ala	Gly	Val	Phe	Ser	Ala
				290					295					300
Gly	Leu	Ala	Pro	Ala	Ala	Phe	Val	Pro	Asn	Pro	Tyr	Ile	Ile	Ser
				305					310					315
Ala	Ala	Pro	Pro	Gly	Thr	Asp	Pro	Tyr	Thr	Ala	Ala	Gly	Leu	Ala

	320		325		330
Ala Ala Ala Thr	Leu Ala Gly Pro Ala Val Val Pro Pro Gln Tyr				
	335		340		345
Tyr Gly Val Pro Trp	Gly Val Tyr Pro Ala Asn Leu Phe Gln Gln				
	350		355		360
Gln Ala Ala Ala	Ala Ala Asn Asn Thr Ala Ser Gln Gln Ala Ala				
	365		370		375
Ser Gln Ala Gln	Pro Gly Gln Gln Gln Val Leu Arg Ala Gly Ala				
	380		385		390
Gly Gln Arg Pro	Leu Thr Pro Asn Gln Gly Gln Gln Gly Gln Gln				
	395		400		405
Ala Glu Ser Leu	Ala Ala Ala Ala Ala Ala Asn Pro Thr Leu Ala				
	410		415		420
Phe Gly Gln Gly	Leu Ala Thr Gly Met Pro Gly Tyr Gln Val Leu				
	425		430		435
Ala Pro Thr Ala	Tyr Tyr Asp Gln Thr Gly Ala Leu Val Val Gly				
	440		445		450
Pro Gly Ala Arg	Thr Gly Leu Gly Ala Pro Val Arg Leu Met Ala				
	455		460		465
Pro Thr Pro Val	Leu Ile Ser Ser Ala Ala Ala Gln Ala Ala Ala				
	470		475		480
Ala Ala Ala Ala	Gly Gly Thr Ala Ser Ser Leu Thr Gly Ser Thr				
	485		490		495
Asn Gly Leu Phe	Arg Pro Ile Gly Thr Gln Pro Pro Gln Gln Gln				
	500		505		510
Gln Gln Gln Pro	Ser Thr Asn Leu Gln Ser Asn Ser Phe Tyr Gly				
	515		520		525
Ser Ser Ser Leu	Thr Asn Ser Ser Gln Ser Ser Ser Leu Phe Ser				
	530		535		540
His Gly Pro Gly	Gln Pro Gly Ser Thr Ser Leu Gly Phe Gly Ser				
	545		550		555
Gly Asn Ser Leu	Gly Ala Ala Ile Gly Ser Ala Leu Ser Gly Phe				
	560		565		570
Gly Ser Ser Gly	Gly Leu Thr Asn Gly Ser Gly Arg Tyr Ile Ser				
	575		580		585
Ala Ala Pro Gly	Ala Glu Ala Lys Tyr Arg Ser Ala Ser Ser Thr				
	590		595		600
Ser Ser Leu Phe	Ser Ser Ser Ser Gln Leu Phe Pro Pro Ser Arg				
	605		610		615
Leu Arg Tyr Asn	Arg Ser Asp Ile Met Pro Ser Gly Arg Ser Arg				
	620		625		630
Leu Leu Glu Asp	Phe Arg Asn Asn Arg Phe Pro Asn Leu Gln Leu				
	635		640		645
Arg Asp Leu Ile	Gly His Ile Val Glu Phe Ser Gln Asp Gln His				
	650		655		660
Gly Ser Arg Phe	Ile Gln Gln Lys Leu Glu Arg Ala Thr Pro Ala				
	665		670		675
Glu Arg Gln Met	Val Phe Asn Glu Ile Leu Gln Ala Ala Tyr Gln				
	680		685		690
Leu Met Thr Asp	Val Phe Gly Asn Tyr Val Ile Gln Lys Phe Phe				
	695		700		705
Glu Phe Gly Ser	Leu Asp Gln Lys Leu Ala Leu Ala Thr Arg Ile				
	710		715		720
Arg Gly His Val	Leu Pro Leu Ala Leu Gln Met Tyr Gly Cys Arg				
	725		730		735
Val Ile Gln Lys	Ala Leu Glu Ser Ile Ser Ser Asp Gln Gln Val				

	740		745		750
Ile Ser Glu Met	Val Lys Glu Leu Asp	Gly His Val Leu Lys Cys			
	755		760		765
Val Lys Asp Gln	Asn Gly Asn His Val	Val Gln Lys Cys Ile Glu			
	770		775		780
Cys Val Gln Pro	Gln Ser Leu Gln Phe	Ile Ile Asp Ala Phe Lys			
	785		790		795
Gly Gln Val Phe	Val Leu Ser Thr His	Pro Tyr Gly Cys Arg Val			
	800		805		810
Ile Gln Arg Ile	Leu Glu His Cys Thr	Ala Glu Gln Thr Leu Pro			
	815		820		825
Ile Leu Glu Glu	Leu His Gln His Thr	Glu Gln Leu Val Gln Asp			
	830		835		840
Gln Tyr Gly Asn	Tyr Val Ile Gln His	Val Leu Glu His Gly Arg			
	845		850		855
Pro Glu Asp Lys	Ser Lys Ile Val Ser	Glu Ile Arg Gly Lys Val			
	860		865		870
Leu Ala Leu Ser	Gln His Lys Phe Ala	Ser Asn Val Val Glu Lys			
	875		880		885
Cys Val Thr His	Ala Ser Arg Ala Glu	Arg Ala Leu Leu Ile Asp			
	890		895		900
Glu Val Cys Cys	Gln Asn Asp Gly Pro	His Ser Ala Leu Tyr Thr			
	905		910		915
Met Met Lys Asp	Gln Tyr Ala Asn Tyr	Val Val Gln Lys Met Ile			
	920		925		930
Asp Met Ala Glu	Pro Ala Gln Arg Lys	Ile Ile Met His Lys Ile			
	935		940		945
Arg Pro His Ile	Thr Thr Leu Arg Lys	Tyr Thr Tyr Gly Lys His			
	950		955		960
Ile Leu Ala Lys	Leu Glu Lys Tyr Tyr	Leu Lys Asn Ser Pro Asp			
	965		970		975
Leu Gly Pro Ile	Gly Gly Pro Pro Asn	Gly Met Leu			
	980		985		

&lt;210&gt; 23

&lt;211&gt; 1013

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1381261CD1

&lt;400&gt; 23

Met Pro Pro Pro	Ala Glu Val Thr Asp	Pro Ser His Ala Pro Ala
1	5	10 15
Val Leu Arg Gln	Leu Asn Glu Gln Arg	Leu Arg Gly Leu Phe Cys
	20	25 30
Asp Val Thr Leu	Ile Ala Gly Asp Thr	Lys Phe Pro Ala His Arg
	35	40 45
Ser Val Leu Ala	Ala Ser Ser Pro Phe	Phe Arg Glu Ala Leu Leu
	50	55 60
Thr Ser Ala Pro	Leu Pro Leu Pro Pro	Ala Thr Gly Gly Ala Ala
	65	70 75
Pro Asn Pro Ala	Thr Thr Thr Ala Ala	Ser Ser Ser Ser Ser Ser
	80	85 90

Ser Ser Ser Ser	Ser Ser Ser Ser Ser	Ser Ala Ser Ser Ser	Ser
95	100	105	
Ser Ser Ser Ser	Ser Ser Pro Pro Pro	Ala Ser Pro Pro Ala	Ser
110	115	120	
Ser Pro Pro Arg	Val Leu Glu Leu Pro	Gly Val Pro Ala Ala	Ala
125	130	135	
Phe Ser Asp Val	Leu Asn Phe Ile Tyr	Ser Ala Arg Leu Ala	Leu
140	145	150	
Pro Gly Gly Gly	Gly Asp Gly Ala Ala	Val Ala Glu Ile Gly	Ala
155	160	165	
Leu Gly Arg Arg	Leu Gly Ile Ser Arg	Leu Gln Gly Leu Gly	Glu
170	175	180	
Gly Gly Asp Ala	Trp Val Pro Pro Thr	Pro Ala Pro Met Ala	Thr
185	190	195	
Ser Gln Pro Glu	Glu Asp Ser Phe Gly	Pro Gly Pro Arg Pro	Ala
200	205	210	
Gly Glu Trp Glu	Gly Asp Arg Ala Glu	Ala Gln Ala Pro Asp	Leu
215	220	225	
Gln Cys Ser Leu	Pro Arg Arg Pro Leu	Pro Cys Pro Gln Cys	Gly
230	235	240	
Lys Ser Phe Ile	His Pro Lys Arg Leu	Gln Thr His Glu Ala	Gln
245	250	255	
Cys Arg Arg Gly	Ala Ser Thr Arg Gly	Ser Thr Gly Leu Gly	Ala
260	265	270	
Gly Gly Ala Gly	Pro Gly Gly Pro Ala	Gly Val Asp Ala Ser	Ala
275	280	285	
Leu Pro Pro Pro	Val Gly Phe Arg Gly	Gly Pro Glu His Val	Val
290	295	300	
Lys Val Val Gly	Gly His Val Leu Tyr	Val Cys Ala Ala Cys	Glu
305	310	315	
Arg Ser Tyr Val	Thr Leu Ser Ser Leu	Lys Arg His Ser Asn	Val
320	325	330	
His Ser Trp Arg	Arg Lys Tyr Pro Cys	Arg Tyr Cys Glu Lys	Val
335	340	345	
Phe Ala Leu Ala	Glu Tyr Arg Thr Lys	His Glu Val Trp His	Thr
350	355	360	
Gly Glu Arg Arg	Tyr Gln Cys Ile Phe	Cys Trp Glu Thr Phe	Val
365	370	375	
Thr Tyr Tyr Asn	Leu Lys Thr His Gln	Arg Ala Phe His Gly	Ile
380	385	390	
Ser Pro Gly Leu	Leu Ala Ser Glu Lys	Thr Pro Asn Gly Gly	Tyr
395	400	405	
Lys Pro Lys Leu	Asn Thr Leu Lys Leu	Tyr Arg Leu Leu Pro	Met
410	415	420	
Arg Ala Ala Lys	Arg Pro Tyr Lys Thr	Tyr Ser Gln Gly Ala	Pro
425	430	435	
Glu Ala Pro Leu	Ser Pro Thr Leu Asn	Thr Pro Ala Pro Val	Ala
440	445	450	
Met Pro Ala Ser	Pro Pro Pro Gly Pro	Pro Pro Ala Pro Glu	Pro
455	460	465	
Gly Pro Pro Pro	Ser Val Ile Thr Phe	Ala His Pro Ala Pro	Ser
470	475	480	
Val Ile Val His	Gly Gly Ser Ser Ser	Gly Gly Gly Gly Ser	Gly
485	490	495	
Thr Ala Ser Thr	Gly Gly Ser Gln Ala	Ala Ser Val Ile Thr	Tyr
500	505	510	



Thr	Ala	Pro	Pro	Arg	Pro	Pro	Lys	Lys	Arg	Glu	Tyr	Pro	Pro	Pro	515	520	525
Pro	Pro	Glu	Pro	Ala	Ala	Thr	Pro	Thr	Ser	Pro	Ala	Thr	Ala	Val	530	535	540
Ser	Pro	Ala	Thr	Ala	Ala	Gly	Pro	Ala	Met	Ala	Thr	Thr	Thr	Glu	545	550	555
Glu	Ala	Lys	Gly	Arg	Asn	Pro	Arg	Ala	Gly	Arg	Thr	Leu	Thr	Tyr	560	565	570
Thr	Ala	Lys	Pro	Val	Gly	Gly	Ile	Gly	Gly	Gly	Gly	Gly	Pro	Pro	575	580	585
Thr	Gly	Ala	Gly	Arg	Gly	Pro	Ser	Gln	Leu	Gln	Ala	Pro	Pro	Pro	590	595	600
Leu	Cys	Gln	Ile	Thr	Val	Arg	Ile	Gly	Glu	Glu	Ala	Ile	Val	Lys	605	610	615
Arg	Arg	Ile	Ser	Glu	Thr	Asp	Leu	Arg	Pro	Gly	Glu	Leu	Ser	Gly	620	625	630
Glu	Glu	Met	Glu	Glu	Ser	Glu	Glu	Asp	Glu	Glu	Glu	Glu	Asp	Glu	635	640	645
Glu	Glu	Glu	Glu	Glu	Asp	Glu	Glu	Glu	Ser	Lys	Ala	Gly	Gly	Glu	650	655	660
Asp	Gln	Leu	Trp	Arg	Pro	Tyr	Tyr	Ser	Tyr	Lys	Pro	Lys	Arg	Lys	665	670	675
Ala	Gly	Ala	Ala	Gly	Gly	Ala	Ser	Val	Gly	Gly	Ser	Gly	Leu	Pro	680	685	690
Arg	Gly	Arg	Arg	Pro	Pro	Arg	Trp	Arg	Gln	Lys	Leu	Glu	Arg	Arg	695	700	705
Ser	Trp	Glu	Glu	Thr	Pro	Ala	Ala	Glu	Ser	Pro	Ala	Gly	Arg	Ala	710	715	720
Arg	Thr	Glu	Arg	Arg	His	Arg	Cys	Gly	Asp	Cys	Ala	Gln	Thr	Phe	725	730	735
Thr	Thr	Leu	Arg	Lys	Leu	Arg	Lys	His	Gln	Glu	Ala	His	Gly	Gly	740	745	750
Gly	Ser	His	Ser	Ser	Arg	Ala	Gly	Arg	Arg	Pro	Ser	Thr	Arg	Phe	755	760	765
Thr	Cys	Pro	His	Cys	Ala	Lys	Val	Cys	Lys	Thr	Ala	Ala	Ala	Leu	770	775	780
Ser	Arg	His	Gly	Gln	Arg	His	Ala	Ala	Glu	Arg	Pro	Gly	Gly	Thr	785	790	795
Pro	Thr	Pro	Val	Ile	Ala	Tyr	Ser	Lys	Gly	Ser	Ala	Gly	Thr	Arg	800	805	810
Pro	Gly	Asp	Val	Lys	Glu	Glu	Ala	Pro	Gln	Glu	Met	Gln	Val	Ser	815	820	825
Ser	Ser	Ser	Gly	Glu	Ala	Gly	Gly	Gly	Ser	Thr	Ala	Ala	Glu	Glu	830	835	840
Ala	Ser	Glu	Thr	Ala	Ser	Leu	Gln	Asp	Pro	Ile	Ile	Ser	Gly	Gly	845	850	855
Glu	Glu	Pro	Pro	Val	Val	Ala	Ser	Gly	Gly	Ser	Tyr	Val	Tyr	Pro	860	865	870
Pro	Val	Gln	Glu	Phe	Pro	Leu	Ala	Leu	Ile	Gly	Gly	Gly	Arg	Glu	875	880	885
Pro	Gly	Gly	Gly	Arg	Gly	Lys	Ser	Gly	Ser	Glu	Gly	Pro	Val	Gly	890	895	900
Ala	Gly	Glu	Gly	Asp	Arg	Met	Glu	Gly	Ile	Gly	Ala	Ala	Lys	Val	905	910	915
Thr	Phe	Tyr	Pro	Glu	Pro	Tyr	Pro	Leu	Val	Tyr	Gly	Pro	Gln	Leu	920	925	930

```

Leu Ala Ala Tyr Pro Tyr Asn Phe Ser Asn Leu Ala Ala Leu Pro
    935                      940                      945
Val Ala Leu Asn Met Val Leu Pro Asp Glu Lys Gly Ala Gly Ala
    950                      955                      960
Leu Pro Phe Leu Pro Gly Val Phe Gly Tyr Ala Val Asn Pro Gln
    965                      970                      975
Ala Ala Pro Pro Ala Pro Pro Thr Pro Pro Pro Thr Leu Pro
    980                      985                      990
Pro Pro Ile Pro Pro Lys Gly Glu Gly Glu Arg Ala Gly Val Glu
    995                      1000                     1005
Arg Thr Gln Lys Gly Asp Val Gly
    1010

```

```

<210> 24
<211> 141
<212> PRT
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 6803876CD1

```

```

<400> 24
Met Ala Glu Ser Asp Trp Asp Thr Val Thr Val Leu Arg Lys Lys
  1          5          10          15
Gly Pro Thr Ala Ala Gln Ala Lys Ser Lys Gln Ala Ile Leu Ala
    20          25          30
Ala Gln Arg Arg Gly Glu Asp Val Glu Thr Ser Lys Lys Trp Ala
    35          40          45
Ala Gly Gln Asn Lys Gln His Ser Ile Thr Lys Asn Thr Ala Lys
    50          55          60
Leu Asp Arg Glu Thr Glu Glu Leu His His Asp Arg Val Thr Leu
    65          70          75
Glu Val Gly Lys Val Ile Gln Gln Gly Arg Gln Ser Lys Gly Leu
    80          85          90
Thr Gln Lys Asp Leu Ala Thr Lys Ile Asn Glu Lys Pro Gln Val
    95          100         105
Ile Ala Asp Tyr Glu Ser Gly Arg Ala Ile Pro Asn Asn Gln Val
   110         115         120
Leu Gly Lys Ile Glu Arg Ala Ile Asp Val Gly Thr Arg Ser Ala
   125         130         135
Arg Val Leu Arg Ala Gln
    140

```

```

<210> 25
<211> 334
<212> PRT
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7506281CD1

```

```

<400> 25
Met Asp Ala Asp Glu Gly Gln Asp Met Ser Gln Val Ser Gly Lys
  1          5          10          15

```

```

Glu Ser Pro Pro Val Ser Asp Thr Pro Asp Glu Gly Asp Glu Pro
    20                      25                      30
Met Pro Ile Pro Glu Asp Leu Ser Thr Thr Ser Gly Gly Gln Gln
    35                      40                      45
Ser Ser Lys Ser Asp Arg Val Val Val Ile Lys Glu Glu Thr Asn
    50                      55                      60
His Ser Glu Met Ala Glu Asp Leu Cys Lys Ile Gly Ser Glu Arg
    65                      70                      75
Ser Leu Val Leu Asp Arg Leu Ala Ser Asn Val Ala Lys Arg Lys
    80                      85                      90
Ser Ser Met Pro Gln Lys Phe Leu Gly Asp Lys Gly Leu Ser Asp
    95                      100                     105
Thr Pro Tyr Asp Ser Ser Ala Ser Tyr Glu Lys Glu Asn Glu Met
   110                      115                     120
Met Lys Ser His Val Met Asp Gln Ala Ile Asn Asn Ala Ile Asn
   125                      130                     135
Tyr Leu Gly Ala Glu Ser Leu Arg Pro Leu Val Gln Thr Pro Pro
   140                      145                     150
Gly Gly Ser Glu Val Val Pro Val Ile Ser Pro Met Tyr Gln Leu
   155                      160                     165
His Lys Pro Leu Ala Glu Gly Thr Pro Arg Ser Asn His Ser Ala
   170                      175                     180
Gln Asp Ser Ala Val Glu Asn Leu Leu Leu Leu Ser Lys Ala Lys
   185                      190                     195
Leu Val Pro Ser Glu Arg Glu Ala Ser Pro Ser Asn Ser Cys Gln
   200                      205                     210
Asp Ser Thr Asp Thr Glu Ser Asn Asn Glu Glu Gln Arg Ser Gly
   215                      220                     225
Leu Ile Tyr Leu Thr Asn His Ile Ala Pro His Ala Arg Asn Gly
   230                      235                     240
Leu Ser Leu Lys Glu Glu His Arg Ala Tyr Asp Leu Leu Arg Ala
   245                      250                     255
Ala Ser Glu Asn Ser Gln Asp Ala Leu Arg Val Val Ser Thr Ser
   260                      265                     270
Gly Glu Gln Met Lys Val Tyr Lys Cys Glu His Cys Arg Val Leu
   275                      280                     285
Phe Leu Asp His Val Met Tyr Thr Ile His Met Gly Cys His Gly
   290                      295                     300
Phe Arg Asp Pro Phe Glu Cys Asn Met Cys Gly Tyr His Ser Gln
   305                      310                     315
Asp Arg Tyr Glu Phe Ser Ser His Ile Thr Arg Gly Glu His Arg
   320                      325                     330
Phe His Met Ser

```

&lt;210&gt; 26

&lt;211&gt; 439

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7506175CD1

&lt;400&gt; 26

Met Asp Thr Ser Asp Leu Phe Ala Ser Cys Arg Lys Gly Asp Val

1	5	10	15
Gly Arg Val Arg Tyr	Leu Leu Glu Gln Arg	Asp Val Glu Val Asn	
20	25	30	
Val Arg Asp Lys Trp	Asp Ser Thr Pro Leu	Tyr Tyr Ala Cys Leu	
35	40	45	
Cys Gly His Glu Glu	Leu Val Leu Tyr Leu	Leu Ala Asn Gly Ala	
50	55	60	
Arg Cys Glu Ala Asn	Thr Phe Asp Gly Glu	Arg Cys Leu Tyr Gly	
65	70	75	
Ala Leu Ser Asp Pro	Ile Arg Arg Ala Leu	Arg Asp Tyr Lys Gln	
80	85	90	
Val Thr Ala Ser Cys	Arg Arg Arg Asp Tyr	Tyr Asp Asp Phe Leu	
95	100	105	
Gln Arg Leu Leu Glu	Gln Gly Ile His Ser	Asp Val Val Phe Val	
110	115	120	
Val His Gly Lys Pro	Phe Arg Val His Arg	Cys Val Leu Gly Ala	
125	130	135	
Arg Ser Ala Tyr Phe	Ala Asn Met Leu Asp	Thr Lys Trp Lys Gly	
140	145	150	
Lys Ser Val Val Val	Leu Arg His Pro Leu	Ile Asn Pro Val Ala	
155	160	165	
Phe Gly Ala Leu Leu	Gln Tyr Leu Tyr Thr	Val Ala Ser Lys Pro	
170	175	180	
Gly Thr Cys Val Lys	Val Leu Thr Ile Glu	Pro Pro Pro Ala Asp	
185	190	195	
Pro Arg Leu Arg Glu	Asp Met Ala Leu Leu	Ala Asp Cys Ala Leu	
200	205	210	
Pro Pro Glu Leu Arg	Gly Asp Leu Trp Glu	Leu Pro Phe Pro Cys	
215	220	225	
Pro Asp Gly Phe Asn	Ser Cys Pro Asp Ile	Cys Phe Arg Val Ala	
230	235	240	
Gly Cys Ser Phe Leu	Cys His Lys Ala Phe	Phe Cys Gly Arg Ser	
245	250	255	
Asp Tyr Phe Arg Ala	Leu Leu Asp Asp His	Phe Arg Glu Ser Glu	
260	265	270	
Glu Pro Ala Thr Ser	Gly Gly Pro Pro Ala	Val Thr Leu His Gly	
275	280	285	
Ile Ser Pro Asp Val	Phe Thr His Val Leu	Tyr Tyr Met Tyr Ser	
290	295	300	
Asp His Thr Glu Leu	Ser Pro Glu Ala Ala	Tyr Asp Val Leu Ser	
305	310	315	
Val Ala Asp Met Tyr	Leu Leu Pro Gly Leu	Lys Arg Leu Cys Gly	
320	325	330	
Arg Ser Leu Ala Gln	Met Leu Asp Glu Asp	Thr Val Val Gly Val	
335	340	345	
Trp Arg Val Ala Lys	Leu Phe Arg Leu Ala	Arg Leu Glu Asp Gln	
350	355	360	
Cys Thr Glu Tyr Met	Ala Lys Val Ile Glu	Lys Leu Val Glu Arg	
365	370	375	
Glu Asp Phe Val Glu	Ala Val Lys Glu Glu	Ala Ala Ala Val Ala	
380	385	390	
Ala Arg Gln Glu Thr	Asp Ser Ile Pro Leu	Val Asp Asp Ile Arg	
395	400	405	
Phe His Val Ala Ser	Thr Val Gln Thr Tyr	Ser Ala Ile Glu Glu	
410	415	420	
Ala Gln Gln Arg Leu	Arg Ala Leu Glu Asp	Leu Leu Val Ser Ile	

	425	430	435
Gly Leu Asp Cys			

```
<210> 27
<211> 448
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7506303CD1
```

<400> 27														
Met	Glu	Thr	Ile	Leu	Glu	Gln	Gln	Arg	Arg	Tyr	His	Glu	Glu	Lys
1				5					10					15
Glu	Arg	Leu	Met	Asp	Val	Met	Ala	Lys	Glu	Met	Leu	Thr	Lys	Lys
				20					25					30
Ser	Thr	Leu	Arg	Asp	Gln	Ile	Asn	Ser	Asp	His	Arg	Thr	Arg	Ala
				35					40					45
Met	Gln	Asp	Ile	Cys	Val	Pro	Met	Ser	Val	Glu	Phe	Glu	Glu	Leu
				50					55					60
Leu	Lys	Ala	Arg	Glu	Asn	Pro	Ser	Glu	Glu	Ala	Gln	Asn	Leu	Val
				65					70					75
Glu	Phe	Thr	Asp	Glu	Glu	Gly	Tyr	Gly	Arg	Tyr	Leu	Asp	Leu	His
				80					85					90
Asp	Cys	Tyr	Leu	Lys	Tyr	Ile	Asn	Leu	Lys	Ala	Ser	Glu	Lys	Leu
				95					100					105
Asp	Tyr	Ile	Thr	Tyr	Leu	Ser	Ile	Phe	Asp	Gln	Leu	Phe	Asp	Ile
				110					115					120
Pro	Lys	Glu	Arg	Lys	Asn	Ala	Glu	Tyr	Lys	Arg	Tyr	Leu	Glu	Met
				125					130					135
Leu	Leu	Glu	Tyr	Leu	Gln	Asp	Tyr	Thr	Asp	Arg	Val	Lys	Pro	Leu
				140					145					150
Gln	Asp	Gln	Asn	Glu	Leu	Phe	Gly	Lys	Ile	Gln	Ala	Glu	Phe	Glu
				155					160					165
Lys	Lys	Trp	Glu	Asn	Gly	Thr	Phe	Pro	Gly	Trp	Pro	Lys	Glu	Thr
				170					175					180
Ser	Ser	Ala	Leu	Thr	His	Ala	Gly	Ala	His	Leu	Asp	Leu	Ser	Ala
				185					190					195
Phe	Ser	Ser	Trp	Glu	Glu	Leu	Ala	Ser	Leu	Gly	Leu	Asp	Arg	Leu
				200					205					210
Lys	Ser	Ala	Leu	Leu	Ala	Leu	Gly	Leu	Lys	Cys	Gly	Gly	Thr	Leu
				215					220					225
Glu	Glu	Arg	Ala	Gln	Arg	Leu	Phe	Ser	Thr	Lys	Gly	Lys	Ser	Leu
				230					235					240
Glu	Ser	Leu	Asp	Thr	Ser	Leu	Phe	Ala	Lys	Asn	Pro	Lys	Ser	Lys
				245					250					255
Gly	Thr	Lys	Arg	Asp	Thr	Glu	Arg	Asn	Lys	Asp	Ile	Ala	Phe	Leu
				260					265					270
Glu	Ala	Gln	Ile	Tyr	Glu	Tyr	Val	Glu	Ile	Leu	Gly	Glu	Gln	Arg
				275					280					285
His	Leu	Thr	His	Glu	Asn	Val	Gln	Arg	Lys	Gln	Ala	Arg	Thr	Gly
				290					295					300
Glu	Glu	Arg	Glu	Glu	Glu	Glu	Glu	Glu	Gln	Ile	Ser	Glu	Ser	Glu
				305					310					315

```

Ser Glu Asp Glu Glu Asn Glu Ile Ile Tyr Asn Pro Lys Asn Leu
      320                      325                      330
Pro Leu Gly Trp Asp Gly Lys Pro Ile Pro Tyr Trp Leu Tyr Lys
      335                      340                      345
Leu His Gly Leu Asn Ile Asn Tyr Asn Cys Glu Ile Cys Gly Asn
      350                      355                      360
Tyr Thr Tyr Arg Gly Pro Lys Ala Phe Gln Arg His Phe Ala Glu
      365                      370                      375
Trp Arg His Ala His Gly Met Arg Cys Leu Gly Ile Pro Asn Thr
      380                      385                      390
Ala His Phe Ala Asn Val Thr Gln Ile Glu Asp Ala Val Ser Leu
      395                      400                      405
Trp Ala Lys Leu Lys Leu Gln Lys Ala Ser Glu Arg Trp Gln Pro
      410                      415                      420
Asp Thr Glu Glu Glu Tyr Glu Asp Ser Ser Gly Asn Val Val Asn
      425                      430                      435
Lys Lys Thr Tyr Glu Asp Leu Lys Arg Gln Gly Leu Leu
      440                      445

```

<210> 28  
 <211> 104  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7353336CD1

```

<400> 28
Met Pro Lys Arg Lys Ala Glu Gly Asp Ala Lys Gly Asp Lys Ala
  1                      5                      10                      15
Lys Val Lys Asp Glu Pro Gln Arg Arg Ser Ala Arg Leu Ser Ala
      20                      25                      30
Lys Pro Ala Pro Pro Lys Pro Glu Pro Lys Pro Lys Lys Ala Pro
      35                      40                      45
Ala Lys Lys Gly Glu Lys Val Pro Lys Gly Lys Lys Gly Lys Ala
      50                      55                      60
Asp Ala Gly Lys Glu Gly Asn Asn Pro Ala Glu Asn Gly Asp Ala
      65                      70                      75
Lys Thr Asp Gln Ala Gln Lys Ala Glu Gly Ala Gly Asp Ala Lys
      80                      85                      90
Thr Asp Gln Ala Gln Lys Ala Glu Gly Ala Gly Asp Ala Lys
      95                      100

```

<210> 29  
 <211> 769  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 3001652CD1

```

<400> 29
Met Gly Pro Pro Leu Ala Pro Arg Pro Ala His Val Pro Gly Glu
  1                      5                      10                      15

```

Ala	Gly	Pro	Arg	Arg	Thr	Arg	Glu	Ser	Arg	Pro	Gly	Ala	Val	Ser
				20					25					30
Phe	Ala	Asp	Val	Ala	Val	Tyr	Phe	Ser	Pro	Glu	Glu	Trp	Glu	Cys
				35					40					45
Leu	Arg	Pro	Ala	Gln	Arg	Ala	Leu	Tyr	Arg	Asp	Val	Met	Arg	Glu
				50					55					60
Thr	Phe	Gly	His	Leu	Gly	Ala	Leu	Gly	Phe	Ser	Val	Pro	Lys	Pro
				65					70					75
Ala	Phe	Ile	Ser	Trp	Val	Glu	Gly	Glu	Val	Glu	Ala	Trp	Ser	Pro
				80					85					90
Glu	Ala	Gln	Asp	Pro	Asp	Gly	Glu	Ser	Ser	Ala	Ala	Phe	Ser	Arg
				95					100					105
Gly	Gln	Gly	Gln	Glu	Ala	Gly	Ser	Arg	Asp	Gly	Asn	Glu	Glu	Lys
				110					115					120
Glu	Arg	Leu	Lys	Lys	Cys	Pro	Lys	Gln	Lys	Glu	Val	Ala	His	Glu
				125					130					135
Val	Ala	Val	Lys	Glu	Trp	Trp	Pro	Ser	Val	Ala	Cys	Pro	Glu	Phe
				140					145					150
Cys	Asn	Pro	Arg	Gln	Ser	Pro	Met	Asn	Pro	Trp	Leu	Lys	Asp	Thr
				155					160					165
Leu	Thr	Arg	Arg	Leu	Pro	His	Ser	Cys	Pro	Asp	Cys	Gly	Arg	Asn
				170					175					180
Phe	Ser	Tyr	Pro	Ser	Leu	Leu	Ala	Ser	His	Gln	Arg	Val	His	Ser
				185					190					195
Gly	Glu	Arg	Pro	Phe	Ser	Cys	Gly	Gln	Cys	Gln	Ala	Arg	Phe	Ser
				200					205					210
Gln	Arg	Arg	Tyr	Leu	Leu	Gln	His	Gln	Phe	Ile	His	Thr	Gly	Glu
				215					220					225
Lys	Pro	Tyr	Pro	Cys	Pro	Asp	Cys	Gly	Arg	Arg	Phe	Arg	Gln	Arg
				230					235					240
Gly	Ser	Leu	Ala	Ile	His	Arg	Arg	Ala	His	Thr	Gly	Glu	Lys	Pro
				245					250					255
Tyr	Ala	Cys	Ser	Asp	Cys	Lys	Ser	Arg	Phe	Thr	Tyr	Pro	Tyr	Leu
				260					265					270
Leu	Ala	Ile	His	Gln	Arg	Lys	His	Thr	Gly	Glu	Lys	Pro	Tyr	Ser
				275					280					285
Cys	Pro	Asp	Cys	Ser	Leu	Arg	Phe	Ala	Tyr	Thr	Ser	Leu	Leu	Ala
				290					295					300
Ile	His	Arg	Arg	Ile	His	Thr	Gly	Glu	Lys	Pro	Tyr	Pro	Cys	Pro
				305					310					315
Asp	Cys	Gly	Arg	Arg	Phe	Thr	Tyr	Ser	Ser	Leu	Leu	Leu	Ser	His
				320					325					330
Arg	Arg	Ile	His	Ser	Asp	Ser	Arg	Pro	Phe	Pro	Cys	Val	Glu	Cys
				335					340					345
Gly	Lys	Gly	Phe	Lys	Arg	Lys	Thr	Ala	Leu	Glu	Ala	His	Arg	Trp
				350					355					360
Ile	His	Arg	Ser	Cys	Ser	Glu	Arg	Arg	Ala	Trp	Gln	Gln	Ala	Val
				365					370					375
Val	Gly	Arg	Ser	Glu	Pro	Ile	Pro	Val	Leu	Gly	Gly	Lys	Asp	Pro
				380					385					390
Pro	Val	His	Phe	Arg	His	Phe	Pro	Asp	Ile	Phe	Gln	Glu	Phe	Cys
				395					400					405
Gln	Gln	Arg	Leu	Gln	Asp	Arg	Gly	Val	Pro	Ser	Asn	Ala	Pro	Pro
				410					415					420
Val	Pro	Gly	Gln	Ser	Pro	Arg	Ser	Phe	Phe	Arg	Asp	Arg	Arg	Gln
				425					430					435

```

Ser Ser Ala Val Ala Tyr Cys Gly Tyr Arg Gly Val Ser Glu Ala
440 445 450
Ser Gly Pro Tyr Ile Phe Leu Glu Gly Lys Lys Pro Leu Leu Tyr
455 460 465
Phe Pro Asp Thr Pro Pro Pro Pro Leu Glu Lys Ala Ala Glu Ala
470 475 480
Ala Leu Phe Lys Gly Lys Trp Asp Asp Glu Ala Arg Glu Met Ala
485 490 495
Pro Pro Pro Ala Pro Leu Leu Ala Pro Arg Pro Gly Glu Thr Arg
500 505 510
Pro Gly Cys Arg Lys Pro Gly Thr Val Ser Phe Ala Asp Val Ala
515 520 525
Val Tyr Phe Ser Pro Glu Glu Trp Gly Cys Leu Arg Pro Ala Gln
530 535 540
Arg Ala Leu Tyr Arg Asp Val Met Gln Glu Thr Tyr Gly His Leu
545 550 555
Gly Ala Leu Gly Phe Pro Gly Pro Lys Pro Ala Leu Ile Ser Trp
560 565 570
Met Glu Gln Glu Ser Glu Ala Trp Ser Pro Ala Ala Gln Asp Pro
575 580 585
Glu Lys Gly Glu Arg Leu Gly Gly Ala Arg Arg Gly Asp Val Pro
590 595 600
Asn Arg Lys Glu Glu Glu Pro Glu Glu Val Pro Arg Ala Lys Gly
605 610 615
Pro Arg Lys Ala Pro Val Lys Glu Ser Pro Glu Val Leu Val Glu
620 625 630
Arg Asn Pro Asp Pro Ala Ile Ser Val Ala Pro Ala Arg Ala Gln
635 640 645
Pro Pro Lys Asn Ala Ala Trp Asp Pro Thr Thr Gly Ala Gln Pro
650 655 660
Pro Ala Pro Ile Pro Ser Met Asp Ala Gln Ala Gly Gln Arg Arg
665 670 675
His Val Cys Thr Asp Cys Gly Arg Arg Phe Thr Tyr Pro Ser Leu
680 685 690
Leu Val Ser His Arg Arg Met His Ser Gly Glu Arg Pro Phe Pro
695 700 705
Cys Pro Glu Cys Gly Met Arg Phe Lys Arg Lys Phe Ala Val Glu
710 715 720
Ala His Gln Trp Ile His Arg Ser Cys Ser Gly Gly Arg Arg Gly
725 730 735
Arg Arg Pro Gly Ile Arg Ala Val Pro Arg Ala Pro Val Arg Gly
740 745 750
Asp Arg Asp Pro Pro Val Leu Phe Arg His Tyr Pro Asp Ile Phe
755 760 765
Glu Glu Cys Gly

```

&lt;210&gt; 30

&lt;211&gt; 1081

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1689128CD1



&lt;400&gt; 30

```

Met Pro Arg Arg Lys Gln Glu Gln Pro Lys Arg Leu Pro Ser His
 1          5          10          15
Val Ser Arg Gln Glu Glu Ala Glu Gly Glu Leu Ser Glu Gly Glu
          20          25          30
His Trp Tyr Gly Asn Ser Ser Glu Thr Pro Ser Glu Ala Ser Tyr
          35          40          45
Gly Glu Val Gln Glu Asn Tyr Lys Leu Ser Leu Glu Asp Arg Ile
          50          55          60
Gln Glu Gln Ser Thr Ser Pro Asp Thr Ser Leu Gly Ser Thr Thr
          65          70          75
Pro Ser Ser His Thr Leu Glu Leu Val Ala Leu Asp Ser Glu Val
          80          85          90
Leu Arg Asp Ser Leu Gln Cys Gln Asp His Leu Ser Pro Gly Val
          95          100          105
Ser Ser Leu Cys Asp Asp Asp Pro Gly Ser Asn Lys Pro Leu Ser
          110          115          120
Ser Asn Leu Arg Arg Leu Leu Glu Ala Gly Ser Leu Lys Leu Asp
          125          130          135
Ala Ala Ala Thr Ala Asn Gly Arg Val Glu Ser Pro Val Asn Val
          140          145          150
Gly Ser Asn Leu Ser Phe Ser Pro Pro Ser His His Ala Gln Gln
          155          160          165
Leu Ser Val Leu Ala Arg Lys Leu Ala Glu Lys Gln Glu Gln Asn
          170          175          180
Asp Gln Tyr Thr Pro Ser Asn Arg Phe Ile Trp Asn Gln Gly Lys
          185          190          195
Trp Leu Pro Asn Ser Thr Thr Thr Cys Ser Leu Ser Pro Asp Ser
          200          205          210
Ala Ile Leu Lys Leu Lys Ala Ala Ala Asn Ala Val Leu Gln Asp
          215          220          225
Lys Ser Leu Thr Arg Thr Glu Glu Thr Met Arg Phe Glu Ser Phe
          230          235          240
Ser Ser Pro Phe Ser Ser Gln Ser Ala Ser Ser Thr Leu Ala Ala
          245          250          255
Leu Ser Lys Lys Val Ser Glu Arg Ser Leu Thr Pro Gly Gln Glu
          260          265          270
His Pro Pro Pro Ala Ser Ser Phe Leu Ser Leu Ala Ser Met Thr
          275          280          285
Ser Ser Ala Ala Leu Leu Lys Glu Val Ala Ala Arg Ala Ala Gly
          290          295          300
Ser Leu Leu Ala Glu Lys Ser Ser Leu Leu Pro Glu Asp Pro Leu
          305          310          315
Pro Pro Pro Pro Ser Glu Lys Lys Pro Glu Lys Val Thr Pro Pro
          320          325          330
Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Gln Ser
          335          340          345
Leu Glu Leu Leu Leu Leu Pro Val Pro Lys Gly Arg Val Ser Lys
          350          355          360
Pro Ser Asn Ser Ala Ser Glu Glu Glu Ser Gly Lys Pro Phe Gln
          365          370          375
Cys Pro Ile Cys Gly Leu Val Ile Lys Arg Lys Ser Tyr Trp Lys
          380          385          390
Arg His Met Val Ile His Thr Gly Leu Lys Ser His Gln Cys Pro
          395          400          405
Leu Cys Pro Phe Arg Cys Ala Arg Lys Asp Asn Leu Lys Ser His

```

	410		415		420
Met Lys Val His	Gln His Gln Asp Arg Gly Glu Thr Phe Gln Cys				
	425		430		435
Gln Leu Cys Pro	Phe Thr Ser Ser Arg His Phe Ser Leu Lys Leu				
	440		445		450
His Met Arg Cys	His Gln His Phe Leu Arg Thr Glu Ala Lys Val				
	455		460		465
Lys Glu Glu Ile	Pro Asp Pro Asp Val Lys Gly Ser Pro His Leu				
	470		475		480
Ser Asp Ser Ala	Cys Leu Gly Gln Gln Arg Glu Gly Gly Gly Thr				
	485		490		495
Glu Leu Val Gly	Thr Met Met Thr Ser Asn Thr Pro Glu Arg Thr				
	500		505		510
Ser Gln Gly Gly	Ala Gly Val Ser Pro Leu Leu Val Lys Glu Glu				
	515		520		525
Pro Lys Glu Asp	Asn Gly Leu Pro Thr Ser Phe Thr Leu Asn Ala				
	530		535		540
Ala Asp Arg Pro	Ala Asn His Thr Lys Leu Lys Asp Pro Ser Glu				
	545		550		555
Tyr Val Ala Asn	Ser Ala Ser Ala Leu Phe Ser Gln Asp Ile Ser				
	560		565		570
Val Lys Met Ala	Ser Asp Phe Leu Met Lys Leu Ser Ala Ala Asn				
	575		580		585
Gln Lys Glu Pro	Met Asn Leu Asn Phe Lys Val Lys Glu Glu Pro				
	590		595		600
Lys Glu Gly Glu	Ser Leu Ser Thr Thr Leu Pro Arg Ser Ser Tyr				
	605		610		615
Val Phe Ser Pro	Glu Ser Glu Val Ser Ala Pro Gly Val Ser Glu				
	620		625		630
Asp Ala Leu Lys	Pro Gln Glu Gly Lys Gly Ser Val Leu Arg Arg				
	635		640		645
Asp Val Ser Val	Lys Ala Ala Ser Glu Leu Leu Met Lys Leu Ser				
	650		655		660
Ala Glu Ser Tyr	Lys Glu Thr Gln Met Val Lys Ile Lys Glu Glu				
	665		670		675
Pro Met Glu Val	Asp Ile Gln Asp Ser His Val Ser Ile Ser Pro				
	680		685		690
Ser Arg Asn Val	Gly Tyr Ser Thr Leu Ile Gly Arg Glu Lys Thr				
	695		700		705
Glu Pro Leu Gln	Lys Met Pro Glu Gly Arg Val Pro Pro Glu Arg				
	710		715		720
Asn Leu Phe Ser	Gln Asp Ile Ser Val Lys Met Ala Ser Glu Leu				
	725		730		735
Leu Phe Gln Leu	Ser Glu Lys Val Ser Lys Glu His Asn His Thr				
	740		745		750
Lys Glu Asn Thr	Ile Arg Thr Thr Thr Ser Pro Phe Phe Ser Glu				
	755		760		765
Asp Thr Phe Arg	Gln Ser Pro Phe Thr Ser Asn Ser Lys Glu Leu				
	770		775		780
Leu Pro Ser Asp	Ser Val Leu His Gly Arg Ile Ser Ala Pro Glu				
	785		790		795
Thr Glu Lys Ile	Val Leu Glu Ala Gly Asn Gly Leu Pro Ser Trp				
	800		805		810
Lys Phe Asn Asp	Gln Leu Phe Pro Cys Asp Val Cys Gly Lys Val				
	815		820		825
Phe Gly Arg Gln	Gln Thr Leu Ser Arg His Leu Ser Leu His Thr				

	830	835	840
Glu Glu Arg Lys Tyr Lys Cys His Leu Cys Pro Tyr Ala Ala Lys			
	845	850	855
Cys Arg Ala Asn Leu Asn Gln His Leu Thr Val His Ser Val Lys			
	860	865	870
Leu Val Ser Thr Asp Thr Glu Asp Ile Val Ser Ala Val Thr Ser			
	875	880	885
Glu Gly Ser Asp Gly Lys Lys His Pro Tyr Tyr Tyr Ser Cys His			
	890	895	900
Val Cys Gly Phe Glu Thr Glu Leu Asn Val Gln Phe Val Ser His			
	905	910	915
Met Ser Leu His Val Asp Lys Glu Gln Trp Met Phe Ser Ile Cys			
	920	925	930
Cys Thr Ala Cys Asp Phe Val Thr Met Glu Glu Ala Glu Ile Lys			
	935	940	945
Thr His Ile Gly Thr Lys His Thr Gly Glu Asp Arg Lys Thr Pro			
	950	955	960
Ser Glu Ser Asn Ser Pro Ser Ser Ser Ser Leu Ser Ala Leu Ser			
	965	970	975
Asp Ser Ala Asn Ser Lys Asp Asp Ser Asp Gly Ser Gln Lys Asn			
	980	985	990
Lys Gly Gly Asn Asn Leu Leu Val Ile Ser Val Met Pro Gly Ser			
	995	1000	1005
Gln Pro Ser Leu Asn Ser Glu Glu Lys Pro Glu Lys Gly Phe Glu			
	1010	1015	1020
Cys Val Phe Cys Asn Phe Val Cys Lys Thr Lys Asn Met Phe Glu			
	1025	1030	1035
Arg His Leu Gln Ile His Leu Ile Thr Arg Met Phe Glu Cys Asp			
	1040	1045	1050
Val Cys His Lys Phe Met Lys Thr Pro Glu Gln Leu Leu Glu His			
	1055	1060	1065
Lys Lys Cys His Thr Val Pro Thr Gly Gly Leu Asn Ser Gly Gln			
	1070	1075	1080
Trp			

&lt;210&gt; 31

&lt;211&gt; 1007

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2362969CD1

&lt;400&gt; 31

Met Glu Leu Asp Leu Asn Ser Ser Ser Glu Asp Asn Lys Pro Gly		
1	5	10
Lys Arg Val Arg Thr Asn Ser Arg Ser Thr Pro Thr Thr Pro Gln		
	20	25
Gly Lys Pro Glu Thr Thr Phe Leu Asp Gln Gly Cys Ser Ser Pro		
	35	40
Val Leu Ile Asp Cys Pro His Pro Asn Cys Asn Lys Lys Tyr Lys		
	50	55
His Ile Asn Gly Leu Arg Tyr His Gln Ala His Ala His Leu Asp		
	65	70
		75

Pro	Glu	Asn	Lys	Leu	Glu	Phe	Glu	Pro	Asp	Ser	Glu	Asp	Lys	Ile	80	85	90
Ser	Asp	Cys	Glu	Glu	Gly	Leu	Ser	Asn	Val	Ala	Leu	Glu	Cys	Ser	95	100	105
Glu	Pro	Ser	Thr	Ser	Val	Ser	Ala	Tyr	Asp	Gln	Leu	Lys	Ala	Pro	110	115	120
Ala	Ser	Pro	Gly	Ala	Gly	Asn	Pro	Pro	Gly	Thr	Pro	Lys	Gly	Lys	125	130	135
Arg	Glu	Leu	Met	Ser	Asn	Gly	Pro	Gly	Ser	Ile	Ile	Gly	Ala	Lys	140	145	150
Ala	Gly	Lys	Asn	Ser	Gly	Lys	Lys	Lys	Gly	Leu	Asn	Asn	Glu	Leu	155	160	165
Asn	Asn	Leu	Pro	Val	Ile	Ser	Asn	Met	Thr	Ala	Ala	Leu	Asp	Ser	170	175	180
Cys	Ser	Ala	Ala	Asp	Gly	Ser	Leu	Ala	Ala	Glu	Met	Pro	Lys	Leu	185	190	195
Glu	Ala	Glu	Gly	Leu	Ile	Asp	Lys	Lys	Asn	Leu	Gly	Asp	Lys	Glu	200	205	210
Lys	Gly	Lys	Lys	Ala	Thr	Asn	Cys	Lys	Thr	Asp	Lys	Asn	Leu	Ser	215	220	225
Lys	Leu	Lys	Ser	Ala	Arg	Pro	Ile	Ala	Pro	Ala	Pro	Ala	Pro	Thr	230	235	240
Pro	Pro	Gln	Leu	Ile	Ala	Ile	Pro	Thr	Ala	Thr	Phe	Thr	Thr	Thr	245	250	255
Thr	Thr	Gly	Thr	Ile	Pro	Gly	Leu	Pro	Ser	Leu	Thr	Thr	Thr	Val	260	265	270
Val	Gln	Ala	Thr	Pro	Lys	Ser	Pro	Pro	Leu	Lys	Pro	Ile	Gln	Pro	275	280	285
Lys	Pro	Thr	Ile	Met	Gly	Glu	Pro	Ile	Thr	Val	Asn	Pro	Ala	Leu	290	295	300
Val	Ser	Leu	Lys	Asp	Lys	Lys	Lys	Lys	Glu	Lys	Arg	Lys	Leu	Lys	305	310	315
Asp	Lys	Glu	Gly	Lys	Glu	Thr	Gly	Ser	Pro	Lys	Met	Asp	Ala	Lys	320	325	330
Leu	Gly	Lys	Leu	Glu	Asp	Ser	Lys	Gly	Ala	Ser	Lys	Asp	Leu	Pro	335	340	345
Gly	His	Phe	Leu	Lys	Asp	His	Leu	Asn	Lys	Asn	Glu	Gly	Leu	Ala	350	355	360
Asn	Gly	Leu	Ser	Glu	Ser	Gln	Glu	Ser	Arg	Met	Ala	Ser	Ile	Lys	365	370	375
Ala	Glu	Ala	Asp	Lys	Val	Tyr	Thr	Phe	Thr	Asp	Asn	Ala	Pro	Ser	380	385	390
Pro	Ser	Ile	Gly	Ser	Ala	Ser	Arg	Leu	Glu	Cys	Ser	Thr	Leu	Val	395	400	405
Asn	Gly	Gln	Ala	Pro	Met	Ala	Pro	Leu	His	Val	Leu	Thr	Gln	Asn	410	415	420
Gly	Ala	Glu	Ser	Ser	Ala	Ala	Lys	Thr	Ser	Ser	Pro	Ala	Tyr	Ser	425	430	435
Asp	Ile	Ser	Asp	Ala	Ala	Asp	Asp	Gly	Gly	Ser	Asp	Ser	Arg	Ser	440	445	450
Glu	Gly	Met	Arg	Ser	Lys	Ala	Ser	Ser	Pro	Ser	Asp	Ile	Ile	Ser	455	460	465
Ser	Lys	Asp	Ser	Val	Val	Lys	Gly	His	Ser	Ser	Thr	Thr	Ala	Gln	470	475	480
Ser	Ser	Gln	Leu	Lys	Glu	Ser	His	Ser	Pro	Tyr	Tyr	His	Ser	Tyr	485	490	495

Asp	Pro	Tyr	Tyr	Ser	Pro	Ser	Tyr	Met	His	Pro	Gly	Gln	Val	Gly
				500					505					510
Ala	Pro	Ala	Ala	Gly	Asn	Ser	Gly	Ser	Thr	Gln	Gly	Met	Lys	Ile
				515					520					525
Lys	Lys	Glu	Ser	Glu	Glu	Asp	Ala	Glu	Lys	Lys	Asp	Lys	Ala	Glu
				530					535					540
Gln	Leu	Asp	Ser	Lys	Lys	Val	Asp	His	Asn	Ser	Ala	Ser	Leu	Gln
				545					550					555
Pro	Gln	His	Gln	Ser	Val	Ile	Thr	Gln	Arg	His	Pro	Ala	Leu	Ala
				560					565					570
Gln	Ser	Leu	Tyr	Tyr	Gly	Gln	Tyr	Ala	Tyr	Gly	Leu	Tyr	Met	Asp
				575					580					585
Gln	Lys	Ser	Leu	Met	Ala	Thr	Ser	Pro	Ala	Tyr	Arg	Gln	Gln	Tyr
				590					595					600
Glu	Lys	Tyr	Tyr	Glu	Asp	Gln	Arg	Leu	Ala	Glu	Gln	Lys	Met	Ala
				605					610					615
Gln	Thr	Gly	Arg	Gly	Asp	Cys	Glu	Arg	Lys	Ser	Glu	Leu	Pro	Leu
				620					625					630
Lys	Glu	Leu	Gly	Lys	Glu	Glu	Thr	Lys	Gln	Lys	Asn	Met	Pro	Ser
				635					640					645
Ala	Thr	Ile	Ser	Lys	Ala	Pro	Ser	Thr	Pro	Glu	Pro	Asn	Lys	Asn
				650					655					660
His	Ser	Lys	Leu	Gly	Pro	Ser	Val	Pro	Asn	Lys	Thr	Glu	Glu	Thr
				665					670					675
Gly	Lys	Ser	Gln	Leu	Leu	Ser	Asn	His	Gln	Gln	Gln	Leu	Gln	Ala
				680					685					690
Asp	Ser	Phe	Lys	Ala	Lys	Gln	Met	Glu	Asn	His	Gln	Leu	Ile	Lys
				695					700					705
Glu	Ala	Val	Glu	Met	Lys	Ser	Val	Met	Asp	Ser	Met	Lys	Gln	Thr
				710					715					720
Gly	Val	Asp	Pro	Thr	Ser	Arg	Phe	Lys	Gln	Asp	Pro	Asp	Ser	Arg
				725					730					735
Thr	Trp	His	His	Tyr	Val	Tyr	Gln	Pro	Lys	Tyr	Leu	Asp	Gln	Gln
				740					745					750
Lys	Ser	Glu	Glu	Leu	Asp	Arg	Glu	Lys	Lys	Leu	Lys	Glu	Asp	Ser
				755					760					765
Pro	Arg	Lys	Thr	Pro	Asn	Lys	Glu	Ser	Gly	Val	Pro	Ser	Leu	Pro
				770					775					780
Val	Ser	Leu	Thr	Ser	Ile	Lys	Glu	Glu	Pro	Lys	Glu	Ala	Lys	His
				785					790					795
Pro	Asp	Ser	Gln	Ser	Met	Glu	Glu	Ser	Lys	Leu	Lys	Asn	Asp	Asp
				800					805					810
Arg	Lys	Thr	Pro	Val	Asn	Trp	Lys	Asp	Ser	Arg	Gly	Thr	Arg	Val
				815					820					825
Ala	Val	Ser	Ser	Pro	Met	Ser	Gln	His	Gln	Ser	Tyr	Ile	Gln	Tyr
				830					835					840
Leu	His	Ala	Tyr	Pro	Tyr	Pro	Gln	Met	Tyr	Asp	Pro	Ser	His	Pro
				845					850					855
Ala	Tyr	Arg	Ala	Val	Ser	Pro	Val	Leu	Met	His	Ser	Tyr	Pro	Gly
				860					865					870
Ala	Tyr	Leu	Ser	Pro	Gly	Phe	His	Tyr	Pro	Val	Tyr	Gly	Lys	Met
				875					880					885
Ser	Gly	Arg	Glu	Glu	Thr	Glu	Lys	Val	Asn	Thr	Ser	Pro	Ser	Val
				890					895					900
Asn	Thr	Lys	Thr	Thr	Thr	Glu	Ser	Lys	Ala	Leu	Asp	Leu	Leu	Gln
				905					910					915

```

Gln His Ala Asn Gln Tyr Arg Ser Lys Ser Pro Ala Pro Val Glu
      920      925      930
Lys Ala Thr Ala Glu Arg Glu Arg Glu Ala Glu Arg Glu Arg Asp
      935      940      945
Arg His Ser Pro Phe Gly Gln Arg His Leu His Thr His His His
      950      955      960
Thr His Val Gly Met Gly Tyr Pro Leu Ile Pro Gly Gln Tyr Asp
      965      970      975
Pro Phe Gln Gly Leu Thr Ser Ala Ala Leu Val Ala Ser Gln Gln
      980      985      990
Val Ala Ala Gln Ala Ser Ala Ser Gly Met Phe Pro Gly Gln Arg
      995      1000     1005
Arg Glu

```

&lt;210&gt; 32

&lt;211&gt; 511

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 4753527CD1

&lt;400&gt; 32

```

Met Ala Ala Ala Ile Thr Arg His Gly Arg Pro Gly Gly Ala Leu
  1      5      10      15
Pro Pro Glu Pro Ser Ala Pro Arg Gln Pro Gly Phe Gly Gly Arg
  20      25      30
Gly Arg Ala Glu Pro Pro Glu Glu Glu Glu Glu Glu Glu Glu
  35      40      45
Glu Glu Glu Ala Glu Ala Glu Ala Val Ala Ala Leu Leu Leu Asn
  50      55      60
Gly Gly Ser Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Val Gly
  65      70      75
Gly Gly Glu Ala Glu Thr Met Ser Glu Pro Ser Pro Glu Ser Ala
  80      85      90
Ser Gln Ala Gly Glu Asp Glu Asp Glu Glu Glu Asp Asp Glu Glu
  95     100     105
Glu Glu Asp Glu Ser Ser Ser Ser Gly Gly Glu Glu Glu Ser
  110     115     120
Ser Ala Glu Ser Leu Val Gly Ser Ser Gly Gly Ser Ser Ser Asp
  125     130     135
Glu Thr Arg Ser Leu Ser Pro Gly Ala Ala Ser Ser Ser Ser Gly
  140     145     150
Asp Gly Asp Gly Lys Glu Gly Leu Glu Glu Pro Lys Gly Pro Arg
  155     160     165
Gly Ser Gln Gly Gly Gly Gly Gly Gly Ser Ser Ser Ser Ser Val
  170     175     180
Val Ser Ser Gly Gly Asp Glu Gly Tyr Gly Thr Gly Gly Gly Gly
  185     190     195
Ser Ser Ala Thr Ser Gly Gly Arg Arg Gly Ser Leu Glu Met Ser
  200     205     210
Ser Asp Gly Glu Pro Leu Ser Arg Met Asp Ser Glu Asp Ser Ile
  215     220     225
Ser Ser Thr Ile Met Asp Val Asp Ser Thr Ile Ser Ser Gly Arg

```

	230		235		240
Ser Thr Pro Ala	Met Met Asn Gly Gln Gly	Ser Thr Thr Ser Ser			
	245		250		255
Ser Lys Asn Ile	Ala Tyr Asn Cys Cys Trp Asp Gln Cys Gln Ala				
	260		265		270
Cys Phe Asn Ser	Ser Pro Asp Leu Ala Asp His Ile Arg Ser Ile				
	275		280		285
His Val Asp Gly	Gln Arg Gly Gly Val Phe Val Cys Leu Trp Lys				
	290		295		300
Gly Cys Lys Val	Tyr Asn Thr Pro Ser Thr Ser Gln Ser Trp Leu				
	305		310		315
Gln Arg His Met	Leu Thr His Ser Gly Asp Lys Pro Phe Lys Cys				
	320		325		330
Val Val Gly Gly	Cys Asn Ala Ser Phe Ala Ser Gln Gly Gly Leu				
	335		340		345
Ala Arg His Val	Pro Thr His Phe Ser Gln Gln Asn Ser Ser Lys				
	350		355		360
Val Ser Ser Gln	Pro Lys Ala Lys Glu Glu Ser Pro Ser Lys Ala				
	365		370		375
Gly Met Asn Lys	Arg Arg Lys Leu Lys Asn Lys Arg Arg Arg Ser				
	380		385		390
Leu Pro Arg Pro	His Asp Phe Phe Asp Ala Gln Thr Leu Asp Ala				
	395		400		405
Ile Arg His Arg	Ala Ile Cys Phe Asn Leu Ser Ala His Ile Glu				
	410		415		420
Ser Leu Gly Lys	Gly His Ser Val Val Phe His Ser Thr Val Ile				
	425		430		435
Ala Lys Arg Lys	Glu Asp Ser Gly Lys Ile Lys Leu Leu Leu His				
	440		445		450
Trp Met Pro Glu	Asp Ile Leu Pro Asp Val Trp Val Asn Glu Ser				
	455		460		465
Glu Arg His Gln	Leu Lys Thr Lys Val Val His Leu Ser Lys Leu				
	470		475		480
Pro Lys Asp Thr	Ala Leu Leu Leu Asp Pro Asn Ile Tyr Arg Thr				
	485		490		495
Met Pro Gln Lys	Arg Leu Lys Arg Phe Asp Ile Leu Asn Phe Pro				
	500		505		510
Arg					

&lt;210&gt; 33

&lt;211&gt; 485

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 6928688CD1

&lt;400&gt; 33

Met Lys Asn Ala	Thr Ile Val Met Ser Val Arg Arg Glu Gln Gly
1	5 10 15
Ser Ser Ser Gly	Glu Gly Ser Leu Ser Phe Glu Asp Val Ala Val
	20 25 30
Gly Phe Thr Arg	Glu Glu Trp Gln Phe Leu Asp Gln Ser Gln Lys
	35 40 45

Val	Leu	Tyr	Lys	Glu	Val	Met	Leu	Glu	Asn	Tyr	Ile	Asn	Leu	Val	50	55	60
Ser	Ile	Gly	Tyr	Arg	Gly	Thr	Lys	Pro	Asp	Ser	Leu	Phe	Lys	Leu	65	70	75
Glu	Gln	Gly	Glu	Pro	Pro	Gly	Ile	Ala	Glu	Gly	Ala	Ala	His	Ser	80	85	90
Gln	Ile	Cys	Pro	Gly	Phe	Val	Ile	Gln	Ser	Arg	Arg	Tyr	Ala	Gly	95	100	105
Lys	Asp	Ser	Asp	Ala	Phe	Gly	Gly	Tyr	Gly	Arg	Ser	Cys	Leu	His	110	115	120
Ile	Lys	Arg	Asp	Lys	Thr	Leu	Thr	Gly	Val	Lys	Tyr	His	Arg	Cys	125	130	135
Val	Lys	Pro	Ser	Ser	Pro	Lys	Ser	Gln	Leu	Asn	Asp	Leu	Gln	Lys	140	145	150
Ile	Cys	Ala	Gly	Gly	Lys	Pro	His	Glu	Cys	Ser	Val	Cys	Gly	Arg	155	160	165
Ala	Phe	Ser	Arg	Lys	Ala	Gln	Leu	Ile	Gln	His	Gln	Arg	Thr	Glu	170	175	180
Arg	Gly	Glu	Lys	Pro	His	Gly	Cys	Gly	Glu	Cys	Gly	Lys	Thr	Phe	185	190	195
Met	Arg	Lys	Ile	Gln	Leu	Thr	Glu	His	Gln	Arg	Thr	His	Thr	Gly	200	205	210
Glu	Lys	Pro	His	Glu	Cys	Ser	Glu	Cys	Gly	Lys	Ala	Phe	Ser	Arg	215	220	225
Lys	Ser	Gln	Leu	Met	Val	His	Gln	Arg	Thr	His	Thr	Gly	Glu	Lys	230	235	240
Pro	Tyr	Arg	Cys	Ser	Glu	Cys	Gly	Lys	Ala	Phe	Ser	Arg	Lys	Cys	245	250	255
Arg	Leu	Asn	Arg	His	Gln	Arg	Ser	His	Thr	Gly	Glu	Lys	Leu	Tyr	260	265	270
Gly	Cys	Ser	Val	Cys	Gly	Lys	Ala	Phe	Ser	Gln	Lys	Ala	Tyr	Leu	275	280	285
Thr	Ala	His	Gln	Arg	Leu	His	Thr	Gly	Asp	Lys	Pro	Tyr	Lys	Cys	290	295	300
Ser	Asp	Cys	Gly	Arg	Thr	Phe	Tyr	Phe	Lys	Ser	Asp	Leu	Thr	Arg	305	310	315
His	Gln	Arg	Ile	His	Thr	Gly	Glu	Lys	Pro	Tyr	Glu	Cys	Ser	Glu	320	325	330
Cys	Glu	Lys	Ala	Phe	Arg	Ser	Lys	Ser	Lys	Leu	Ile	Gln	His	Gln	335	340	345
Arg	Thr	His	Thr	Gly	Glu	Arg	Pro	Tyr	Ser	Cys	Arg	Glu	Cys	Gly	350	355	360
Lys	Ala	Phe	Ala	His	Met	Ser	Val	Leu	Ile	Lys	His	Glu	Lys	Thr	365	370	375
His	Ile	Arg	Glu	Thr	Ala	Ile	Asn	Ser	Leu	Thr	Val	Glu	Lys	Pro	380	385	390
Ser	Ser	Arg	Ser	His	Thr	Ser	Leu	Tyr	Met	Ser	Glu	Leu	Ile	Gln	395	400	405
Glu	Gln	Lys	Thr	Val	Asn	Thr	Val	Pro	Ile	Glu	Met	Pro	Ser	Ser	410	415	420
Gly	Thr	Pro	Pro	Leu	Leu	Asn	Lys	Ser	Glu	Arg	Leu	Val	Gly	Arg	425	430	435
Asn	Val	Val	Ile	Val	Glu	Gln	Pro	Phe	Pro	Arg	Asn	Gln	Ala	Phe	440	445	450
Val	Val	Asn	Gln	Glu	Phe	Glu	Gln	Arg	Ile	Ser	Leu	Thr	Asn	Glu	455	460	465



Val Asn Val Ala Pro Ser Val Ile Asn Tyr Ile Leu Tyr Leu Thr  
 470 475 480  
 Asp Ile Val Ser Glu  
 485

<210> 34

<211> 1011

<212> PRT

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 7506388CD1

<400> 34

Met Val Ser Ser Asp Arg Pro Val Ser Leu Glu Asp Glu Val Ser  
 1 5 10 15  
 His Ser Met Lys Glu Met Ile Gly Gly Cys Val Cys Ser Asp  
 20 25 30  
 Glu Arg Gly Trp Ala Glu Asn Pro Leu Val Tyr Cys Asp Gly His  
 35 40 45  
 Gly Cys Ser Val Ala Val His Gln Ala Cys Tyr Gly Ile Val Gln  
 50 55 60  
 Val Pro Thr Gly Pro Trp Phe Cys Arg Lys Cys Glu Ser Gln Glu  
 65 70 75  
 Arg Ala Ala Arg Val Arg Cys Glu Leu Cys Pro His Lys Asp Gly  
 80 85 90  
 Ala Leu Lys Arg Thr Asp Asn Gly Gly Trp Ala His Val Val Cys  
 95 100 105  
 Ala Leu Tyr Ile Pro Glu Val Gln Phe Ala Asn Val Ser Thr Met  
 110 115 120  
 Glu Pro Ile Val Leu Gln Ser Val Pro His Asp Arg Tyr Asn Lys  
 125 130 135  
 Thr Cys Tyr Ile Cys Asp Glu Gln Gly Arg Glu Ser Lys Ala Ala  
 140 145 150  
 Thr Gly Ala Cys Met Thr Cys Asn Lys His Gly Cys Arg Gln Ala  
 155 160 165  
 Phe His Val Thr Cys Ala Gln Phe Ala Gly Leu Leu Cys Glu Glu  
 170 175 180  
 Glu Gly Asn Gly Ala Asp Asn Val Gln Tyr Cys Gly Tyr Cys Lys  
 185 190 195  
 Tyr His Phe Ser Lys Leu Lys Lys Ser Lys Arg Gly Ser Asn Arg  
 200 205 210  
 Ser Tyr Asp Gln Ser Leu Ser Asp Ser Ser Ser His Ser Gln Asp  
 215 220 225  
 Lys His His Glu Lys Glu Lys Lys Lys Tyr Lys Glu Lys Asp Lys  
 230 235 240  
 His Lys Gln Lys His Lys Lys Gln Pro Glu Pro Ser Pro Ala Leu  
 245 250 255  
 Val Pro Ser Leu Thr Val Thr Thr Glu Lys Thr Tyr Thr Ser Thr  
 260 265 270  
 Ser Asn Asn Ser Ile Ser Gly Ser Leu Lys Arg Leu Glu Asp Thr  
 275 280 285  
 Thr Ala Arg Phe Thr Asn Ala Asn Phe Gln Glu Val Ser Ala His  
 290 295 300  
 Thr Ser Ser Gly Lys Asp Val Ser Glu Thr Arg Gly Ser Glu Gly

	305		310		315
Lys Gly Lys Lys	Ser Ser Ala His Ser	Ser Gly Gln Arg Gly Arg			
	320		325		330
Lys Pro Gly Gly	Gly Arg Asn Pro Gly	Thr Thr Val Ser Ala Ala			
	335		340		345
Ser Pro Phe Pro	Gln Gly Ser Phe Ser	Gly Thr Pro Gly Ser Val			
	350		355		360
Lys Ser Ser Ser	Gly Ser Ser Val Gln	Ser Pro Gln Asp Phe Leu			
	365		370		375
Ser Phe Thr Asp	Ser Asp Leu Arg Asn	Asp Ser Tyr Ser His Ser			
	380		385		390
Gln Gln Ser Ser	Ala Thr Lys Asp Val	His Lys Gly Glu Ser Gly			
	395		400		405
Ser Gln Glu Gly	Gly Val Asn Ser Phe	Ser Thr Leu Ile Gly Leu			
	410		415		420
Pro Ser Thr Ser	Ala Val Thr Ser Gln	Pro Lys Ser Phe Glu Asn			
	425		430		435
Ser Pro Gly Asp	Leu Gly Asn Ser Ser	Leu Pro Thr Ala Gly Tyr			
	440		445		450
Lys Arg Ala Gln	Thr Ser Gly Ile Glu	Glu Glu Thr Val Lys Glu			
	455		460		465
Lys Lys Arg Lys	Gly Asn Lys Gln Ser	Lys His Gly Pro Gly Arg			
	470		475		480
Pro Lys Gly Asn	Lys Asn Gln Glu Asn	Val Ser His Leu Ser Val			
	485		490		495
Ser Ser Ala Ser	Pro Thr Ser Ser Val	Ala Ser Ala Ala Gly Ser			
	500		505		510
Ile Thr Ser Ser	Ser Leu Gln Lys Ser	Pro Thr Leu Leu Arg Asn			
	515		520		525
Gly Ser Leu Gln	Ser Leu Ser Val Gly	Ser Ser Pro Val Gly Ser			
	530		535		540
Glu Ile Ser Met	Gln Tyr Arg His Asp	Gly Ala Cys Pro Thr Thr			
	545		550		555
Thr Phe Ser Glu	Leu Leu Asn Ala Ile	His Asn Gly Ile Tyr Asn			
	560		565		570
Ser Asn Asp Val	Ala Val Ser Phe Pro	Asn Val Val Ser Gly Ser			
	575		580		585
Gly Ser Ser Thr	Pro Val Ser Ser Ser	His Leu Pro Gln Gln Ser			
	590		595		600
Ser Gly His Leu	Gln Gln Val Gly Ala	Leu Ser Pro Ser Ala Val			
	605		610		615
Ser Ser Ala Ala	Pro Ala Val Ala Thr	Thr Gln Ala Asn Thr Leu			
	620		625		630
Ser Gly Ser Ser	Leu Ser Gln Ala Pro	Ser His Met Tyr Gly Asn			
	635		640		645
Arg Ser Asn Ser	Ser Met Ala Ala Leu	Ile Ala Gln Ser Glu Asn			
	650		655		660
Asn Gln Thr Asp	Gln Asp Leu Gly Asp	Asn Ser Arg Asn Leu Val			
	665		670		675
Gly Arg Gly Ser	Ser Pro Arg Gly Ser	Leu Ser Pro Arg Ser Pro			
	680		685		690
Val Ser Ser Leu	Gln Ile Arg Tyr Asp	Gln Pro Gly Asn Ser Ser			
	695		700		705
Leu Glu Asn Leu	Pro Pro Val Ala Ala	Ser Ile Glu Gln Leu Leu			
	710		715		720
Glu Arg Gln Trp	Ser Glu Gly Gln Gln	Phe Leu Leu Glu Gln Gly			

	725		730		735
Thr Pro Ser Asp	Ile Leu Gly Met Leu	Lys Ser Leu His Gln Leu			
	740		745		750
Gln Val Glu Asn Arg Arg Leu Glu Glu	Gln Ile Lys Asn Leu Thr				
	755		760		765
Ala Lys Lys Glu Arg Leu Gln Leu Leu	Asn Ala Gln Leu Ser Val				
	770		775		780
Pro Phe Pro Thr Ile Thr Ala Asn Pro	Ser Pro Ser His Gln Ile				
	785		790		795
His Thr Phe Ser Ala Gln Thr Ala Pro	Thr Thr Asp Ser Leu Asn				
	800		805		810
Ser Ser Lys Ser Pro His Ile Gly Asn	Ser Phe Leu Pro Asp Asn				
	815		820		825
Ser Leu Pro Val Leu Asn Gln Asp Leu	Thr Ser Ser Gly Gln Ser				
	830		835		840
Thr Ser Ser Ser Ser Ala Leu Ser Thr	Pro Pro Pro Ala Gly Gln				
	845		850		855
Ser Pro Ala Gln Gln Gly Ser Gly Val	Ser Gly Val Gln Gln Val				
	860		865		870
Asn Gly Val Thr Val Gly Ala Leu Ala	Ser Gly Met Gln Pro Val				
	875		880		885
Thr Ser Thr Ile Pro Ala Val Ser Ala	Val Gly Gly Ile Ile Gly				
	890		895		900
Ala Leu Pro Gly Asn Gln Leu Ala Ile	Asn Gly Ile Val Gly Ala				
	905		910		915
Leu Asn Gly Val Met Gln Thr Pro Val	Thr Met Ser Gln Asn Pro				
	920		925		930
Thr Pro Leu Thr His Thr Thr Val Pro	Pro Asn Ala Thr His Pro				
	935		940		945
Met Pro Ala Thr Leu Thr Asn Ser Ala	Ser Gly Leu Gly Leu Leu				
	950		955		960
Ser Asp Gln Gln Arg Gln Ile Leu Ile	His Gln Gln Gln Phe Gln				
	965		970		975
Gln Leu Leu Asn Ser Gln Gln Leu Thr	Pro Val His Arg His Pro				
	980		985		990
His Phe Thr Gln Leu Pro Pro Thr His	Phe Ser Pro Ser Met Glu				
	995		1000		1005
Ile Met Gln Val Arg Lys					
	1010				

&lt;210&gt; 35

&lt;211&gt; 1675

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7376372CD1

&lt;400&gt; 35

Met Thr Ser Ser Ser	Pro Ala Gly Leu Glu Gly Ser Asp Leu Ser
1	5 10 15
Ser Ile Asn Thr Met Met Ser Ala Val	Met Ser Val Gly Lys Val
20	25 30
Thr Glu Asn Gly Gly Ser Pro Gln Gly	Ile Lys Ser Pro Ser Lys
35	40 45

Pro	Pro	Gly	Pro	Asn	Arg	Ile	Gly	Arg	Arg	Asn	Gln	Glu	Thr	Lys
				50					55					60
Glu	Glu	Lys	Ser	Ser	Tyr	Asn	Cys	Pro	Leu	Cys	Glu	Lys	Ile	Cys
				65					70					75
Thr	Thr	Gln	His	Gln	Leu	Thr	Met	His	Ile	Arg	Gln	His	Asn	Thr
				80					85					90
Asp	Thr	Gly	Gly	Ala	Asp	His	Ser	Cys	Ser	Ile	Cys	Gly	Lys	Ser
				95					100					105
Leu	Ser	Ser	Ala	Ser	Ser	Leu	Asp	Arg	His	Met	Leu	Val	His	Ser
				110					115					120
Gly	Glu	Arg	Pro	Tyr	Lys	Cys	Thr	Val	Cys	Gly	Gln	Ser	Phe	Thr
				125					130					135
Thr	Asn	Gly	Asn	Met	His	Arg	His	Met	Lys	Ile	His	Glu	Lys	Asp
				140					145					150
Pro	Asn	Ser	Ala	Thr	Ala	Thr	Ala	Pro	Pro	Ser	Pro	Leu	Lys	Arg
				155					160					165
Arg	Arg	Leu	Ser	Ser	Lys	Arg	Lys	Leu	Ser	His	Asp	Ala	Glu	Ser
				170					175					180
Glu	Arg	Glu	Asp	Pro	Ala	Pro	Ala	Lys	Lys	Met	Val	Glu	Asp	Gly
				185					190					195
Gln	Ser	Gly	Asp	Leu	Glu	Lys	Lys	Ala	Asp	Glu	Val	Phe	His	Cys
				200					205					210
Pro	Val	Cys	Phe	Lys	Glu	Phe	Val	Cys	Lys	Tyr	Gly	Leu	Glu	Thr
				215					220					225
His	Met	Glu	Thr	His	Ser	Asp	Asn	Pro	Leu	Arg	Cys	Asp	Ile	Cys
				230					235					240
Cys	Val	Thr	Phe	Arg	Thr	His	Arg	Gly	Leu	Leu	Arg	His	Asn	Ala
				245					250					255
Leu	Val	His	Lys	Gln	Leu	Pro	Arg	Asp	Ala	Met	Gly	Arg	Pro	Phe
				260					265					270
Ile	Gln	Asn	Asn	Pro	Ser	Ile	Pro	Ala	Gly	Phe	His	Asp	Leu	Gly
				275					280					285
Phe	Thr	Asp	Phe	Ser	Cys	Arg	Lys	Phe	Pro	Arg	Ile	Ser	Gln	Ala
				290					295					300
Trp	Cys	Glu	Thr	Asn	Leu	Arg	Arg	Cys	Ile	Ser	Glu	Gln	His	Arg
				305					310					315
Phe	Val	Cys	Asp	Thr	Cys	Asp	Lys	Ala	Phe	Pro	Met	Leu	Cys	Ser
				320					325					330
Val	Ala	Leu	His	Lys	Gln	Thr	His	Val	Ala	Ala	Asp	Gln	Gly	Gln
				335					340					345
Glu	Lys	Pro	Gln	Ala	Thr	Pro	Leu	Pro	Gly	Asp	Ala	Leu	Asp	Gln
				350					355					360
Lys	Gly	Phe	Leu	Ala	Leu	Leu	Gly	Leu	Gln	His	Thr	Lys	Asp	Val
				365					370					375
Arg	Pro	Ala	Pro	Ala	Glu	Glu	Pro	Leu	Pro	Asp	Asp	Asn	Gln	Ala
				380					385					390
Ile	Gln	Leu	Gln	Thr	Leu	Lys	Cys	Gln	Leu	Pro	Gln	Asp	Pro	Gly
				395					400					405
Cys	Thr	Asn	Leu	Leu	Ser	Leu	Ser	Pro	Phe	Glu	Ala	Ala	Ser	Leu
				410					415					420
Gly	Gly	Ser	Leu	Thr	Val	Leu	Pro	Ala	Thr	Lys	Asp	Ser	Ile	Lys
				425					430					435
His	Leu	Ser	Leu	Gln	Pro	Phe	Gln	Lys	Gly	Phe	Ile	Ile	Gln	Pro
				440					445					450
Asp	Ser	Ser	Ile	Val	Val	Lys	Pro	Ile	Ser	Gly	Glu	Ser	Ala	Ile
				455					460					465

Glu Leu Ala Asp	Ile Gln Gln Ile Leu Lys Met Ala Ala Ser Ala	470	475	480
Pro Pro Gln Ile	Ser Leu Pro Pro Phe Ser Lys Ala Pro Ala Ala	485	490	495
Pro Leu Gln Ala	Ile Phe Lys His Met Pro Pro Leu Lys Ala Lys	500	505	510
Pro Leu Val Thr	Pro Arg Thr Val Val Gly His Leu His Ala Pro	515	520	525
Ala Ser His Gln	Arg Gln Gln Ala Ser Arg Leu Tyr Gln Pro Gln	530	535	540
Pro Ala Ala Thr	Ala Pro Glu Ala Pro Gln Arg Leu Thr Gly Gly	545	550	555
Gly Leu Gln Arg	Pro Pro Ala Ala Val Gln Val Arg Asp Pro Ala	560	565	570
Pro Arg Gly His	Ala Ala Leu Pro Ala Ala Gln Pro Arg Ala Glu	575	580	585
Leu Pro Gly Gln	Pro Glu Met Lys Thr Gln Leu Glu Gln Asp Ser	590	595	600
Ile Ile Glu Ala	Leu Leu Pro Leu Ser Met Glu Ala Lys Ile Lys	605	610	615
Gln Glu Ile Thr	Glu Gly Glu Leu Lys Ala Phe Met Thr Ala Pro	620	625	630
Gly Gly Lys Lys	Thr Pro Ala Met Arg Lys Val Leu Tyr Pro Cys	635	640	645
Arg Phe Cys Asn	Gln Val Phe Ala Phe Ser Gly Val Leu Arg Ala	650	655	660
His Val Arg Ser	His Leu Gly Ile Ser Pro Tyr Gln Cys Asn Ile	665	670	675
Cys Asp Tyr Ile	Ala Ala Asp Lys Ala Ala Leu Ile Arg His Leu	680	685	690
Arg Thr His Ser	Gly Glu Arg Pro Tyr Ile Cys Lys Ile Cys His	695	700	705
Tyr Pro Phe Thr	Val Lys Ala Asn Cys Glu Arg His Leu Arg Lys	710	715	720
Lys His Leu Lys	Ala Thr Arg Lys Asp Ile Glu Lys Asn Ile Glu	725	730	735
Tyr Val Ser Ser	Ser Ala Ala Glu Leu Val Asp Ala Phe Cys Ala	740	745	750
Pro Asp Thr Val	Cys Arg Leu Cys Gly Glu Asp Leu Lys His Tyr	755	760	765
Arg Ala Leu Arg	Ile His Met Pro Thr His Cys Gly Arg Gly Leu	770	775	780
Gly Gly Gly His	Lys Gly Arg Lys Pro Phe Glu Cys Lys Glu Cys	785	790	795
Ser Ala Ala Phe	Ala Ala Lys Arg Asn Cys Ile His His Ile Leu	800	805	810
Lys Gln His Leu	His Val Pro Glu Gln Asp Ile Glu Ser Tyr Val	815	820	825
Leu Ala Ala Asp	Gly Leu Gly Pro Ala Glu Ala Pro Ala Ala Glu	830	835	840
Ala Ser Gly Arg	Gly Glu Asp Ser Gly Cys Ala Ala Leu Gly Asp	845	850	855
Cys Lys Pro Leu	Thr Ala Phe Leu Glu Pro Gln Asn Gly Phe Leu	860	865	870
His Arg Gly Pro	Thr Gln Pro Pro Pro Pro His Val Ser Ile Lys	875	880	885

Leu	Glu	Pro	Ala	Ser	Ser	Phe	Ala	Val	Asp	Phe	Asn	Glu	Pro	Leu	890	895	900
Asp	Phe	Ser	Gln	Lys	Gly	Leu	Ala	Leu	Val	Gln	Val	Lys	Gln	Glu	905	910	915
Asn	Ile	Ser	Phe	Leu	Ser	Pro	Ser	Ser	Leu	Val	Pro	Tyr	Asp	Cys	920	925	930
Ser	Met	Glu	Pro	Ile	Asp	Leu	Ser	Ile	Pro	Lys	Asn	Phe	Arg	Lys	935	940	945
Gly	Asp	Lys	Asp	Leu	Ala	Thr	Pro	Ser	Glu	Ala	Lys	Lys	Pro	Glu	950	955	960
Glu	Glu	Ala	Gly	Ser	Ser	Glu	Gln	Pro	Ser	Pro	Cys	Pro	Ala	Pro	965	970	975
Gly	Pro	Ser	Leu	Pro	Val	Thr	Leu	Gly	Pro	Ser	Gly	Ile	Leu	Glu	980	985	990
Ser	Pro	Met	Ala	Pro	Ala	Pro	Ala	Ala	Thr	Pro	Glu	Pro	Pro	Ala	995	1000	1005
Gln	Pro	Leu	Gln	Gly	Pro	Val	Gln	Leu	Ala	Val	Pro	Ile	Tyr	Ser	1010	1015	1020
Ser	Ala	Leu	Val	Ser	Ser	Pro	Pro	Leu	Val	Gly	Ser	Ser	Ala	Leu	1025	1030	1035
Leu	Ser	Gly	Thr	Ala	Leu	Leu	Arg	Pro	Leu	Arg	Pro	Lys	Pro	Pro	1040	1045	1050
Leu	Leu	Leu	Pro	Lys	Pro	Pro	Val	Thr	Glu	Glu	Leu	Pro	Pro	Leu	1055	1060	1065
Ala	Ser	Ile	Ala	Gln	Ile	Ile	Ser	Ser	Val	Ser	Ser	Ala	Pro	Thr	1070	1075	1080
Leu	Leu	Lys	Thr	Lys	Val	Ala	Asp	Pro	Gly	Pro	Ala	Ser	Thr	Gly	1085	1090	1095
Ser	Asn	Thr	Thr	Ala	Ser	Asp	Ser	Leu	Gly	Gly	Ser	Val	Pro	Lys	1100	1105	1110
Ala	Ala	Thr	Thr	Ala	Thr	Pro	Ala	Ala	Thr	Thr	Ser	Pro	Lys	Glu	1115	1120	1125
Ser	Ser	Glu	Pro	Pro	Ala	Pro	Ala	Thr	Gly	Pro	Glu	Ala	Ala	Ser	1130	1135	1140
Pro	Thr	Glu	Gln	Gly	Pro	Ala	Gly	Thr	Ser	Lys	Lys	Arg	Gly	Arg	1145	1150	1155
Lys	Arg	Gly	Met	Arg	Ser	Arg	Pro	Arg	Ala	Asn	Ser	Gly	Gly	Val	1160	1165	1170
Asp	Leu	Tyr	Ser	Ser	Gly	Glu	Phe	Ala	Ser	Ile	Glu	Lys	Met	Leu	1175	1180	1185
Ala	Thr	Thr	Asp	Thr	Asn	Lys	Phe	Ser	Pro	Phe	Leu	Gln	Thr	Ala	1190	1195	1200
Glu	Asp	Asn	Thr	Gln	Asp	Glu	Val	Ala	Gly	Ala	Pro	Ala	Asp	His	1205	1210	1215
His	Gly	Pro	Ser	Asp	Glu	Glu	Gln	Gly	Ser	Pro	Pro	Glu	Asp	Lys	1220	1225	1230
Leu	Leu	Arg	Ala	Lys	Arg	Asn	Ser	Tyr	Thr	Asn	Cys	Leu	Gln	Lys	1235	1240	1245
Ile	Thr	Cys	Pro	His	Cys	Pro	Arg	Val	Phe	Pro	Trp	Ala	Ser	Ser	1250	1255	1260
Leu	Gln	Arg	His	Met	Leu	Thr	His	Thr	Asp	Ser	Gln	Ser	Asp	Ala	1265	1270	1275
Gly	Asp	Cys	Ser	Arg	Arg	Gly	Arg	Ser	Gly	Tyr	Asp	Leu	Thr	Ser	1280	1285	1290
Arg	Asp	Arg	Glu	Gln	Pro	Ser	Glu	Gly	Ala	Thr	Glu	Leu	Arg	Gln	1295	1300	1305

```

Val Ala Gly Asp Ala Pro Val Glu Gln Ala Thr Ala Glu Thr Ala
1310                      1315                      1320
Ser Arg Cys Thr Gly Lys Ser Thr Gly Val Gly Arg Ala Met Ser
1325                      1330                      1335
Arg Arg Arg Ser Met Ala Leu Arg Arg Ala Leu Gly Thr Pro Thr
1340                      1345                      1350
Arg Lys Arg Thr Arg Arg Ala Thr Arg Ala Trp Thr Trp Thr Ser
1355                      1360                      1365
Pro Pro Ser Ser Trp Thr Ser Ser Trp Arg Arg Ala Thr Arg Arg
1370                      1375                      1380
Gln Ala Gly Gly Ala Ala Ser Gln Glu Gln Lys Leu Ala Cys Asp
1385                      1390                      1395
Ala Cys Gly Lys Ser Phe Lys Phe Leu Gly Thr Leu Ser Arg His
1400                      1405                      1410
Arg Lys Ala His Gly Arg Pro Gly Gly Pro Arg Thr Arg Arg Glu
1415                      1420                      1425
Met Ala Pro Ala Leu Gln Arg Arg Gly Pro Ser Pro Pro Leu Asn
1430                      1435                      1440
Arg Arg Arg Ser Pro Pro Arg Pro Arg Gln Arg Trp Trp Ser Arg
1445                      1450                      1455
Pro Gly Ser Gly Arg Pro Arg Pro Glu Lys Leu Ala Glu Glu Thr
1460                      1465                      1470
Glu Gly Pro Ser Asp Gly Glu Ser Ala Ala Glu Lys Arg Ser Ser
1475                      1480                      1485
Glu Lys Ser Asp Asp Asp Lys Lys Pro Lys Thr Asp Ser Pro Lys
1490                      1495                      1500
Ser Val Ala Ser Lys Ala Asp Lys Arg Lys Lys Val Cys Ser Val
1505                      1510                      1515
Cys Asn Lys Arg Phe Trp Ser Leu Gln Asp Leu Ser Arg His Met
1520                      1525                      1530
Arg Ser His Thr Gly Glu Arg Pro Tyr Lys Cys Gln Thr Cys Glu
1535                      1540                      1545
Arg Thr Phe Thr Leu Lys His Ser Leu Val Arg His Gln Arg Ile
1550                      1555                      1560
His Gln Lys Ala Arg His Ala Lys His His Gly Lys Asp Ser Asp
1565                      1570                      1575
Lys Glu Glu Arg Gly Glu Glu Asp Ser Glu Asn Glu Ser Thr His
1580                      1585                      1590
Ser Gly Asn Asn Ala Val Ser Glu Asn Glu Ala Glu Leu Ala Pro
1595                      1600                      1605
Asn Ala Ser Asn His Met Ala Val Thr Arg Ser Arg Lys Glu Gly
1610                      1615                      1620
Leu Ala Ser Ala Thr Lys Asp Cys Ser His Arg Glu Glu Lys Val
1625                      1630                      1635
Thr Gln Gly Gly Arg Leu Ser Leu Ala Arg Val Thr Leu Thr Gln
1640                      1645                      1650
Arg Ala Arg Arg Pro Trp Gly Arg Thr Cys Trp Ser Arg Ala Ala
1655                      1660                      1665
Arg Gly Leu Pro Thr Gln Ser Trp His Ser
1670                      1675

```

&lt;210&gt; 36

&lt;211&gt; 724

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2754344CD1

&lt;400&gt; 36

Met	Pro	Val	Arg	Phe	Lys	Gly	Leu	Ser	Glu	Tyr	Gln	Arg	Asn	Phe
1				5					10					15
Leu	Trp	Lys	Lys	Ser	Tyr	Leu	Ser	Glu	Ser	Cys	Asn	Ser	Ser	Val
				20					25					30
Gly	Arg	Lys	Tyr	Pro	Trp	Ala	Gly	Leu	Arg	Ser	Asp	Gln	Leu	Gly
				35					40					45
Ile	Thr	Lys	Glu	Pro	Ser	Phe	Ile	Ser	Lys	Arg	Arg	Val	Pro	Tyr
				50					55					60
His	Asp	Pro	Gln	Ile	Ser	Lys	Ser	Leu	Glu	Trp	Asn	Gly	Ala	Ile
				65					70					75
Ser	Glu	Ser	Asn	Val	Val	Ala	Ser	Pro	Glu	Pro	Glu	Ala	Pro	Glu
				80					85					90
Thr	Pro	Lys	Ser	Gln	Glu	Ala	Glu	Gln	Lys	Asp	Val	Ile	Gln	Glu
				95					100					105
Arg	Val	His	Ser	Leu	Glu	Ala	Ser	Arg	Val	Pro	Lys	Arg	Thr	Arg
				110					115					120
Ser	His	Ser	Ala	Asp	Ser	Arg	Ala	Glu	Gly	Ala	Ser	Asp	Val	Glu
				125					130					135
Asn	Asn	Glu	Gly	Val	Thr	Asn	His	Thr	Pro	Val	Asn	Glu	Asn	Val
				140					145					150
Glu	Leu	Glu	His	Ser	Thr	Lys	Val	Leu	Ser	Glu	Asn	Val	Asp	Asn
				155					160					165
Gly	Leu	Asp	Arg	Leu	Leu	Arg	Lys	Lys	Ala	Gly	Leu	Thr	Val	Val
				170					175					180
Pro	Ser	Tyr	Asn	Ala	Leu	Arg	Asn	Ser	Glu	Tyr	Gln	Arg	Gln	Phe
				185					190					195
Val	Trp	Lys	Thr	Ser	Lys	Glu	Thr	Ala	Pro	Ala	Phe	Ala	Ala	Asn
				200					205					210
Gln	Val	Phe	His	Asn	Lys	Ser	Gln	Phe	Val	Pro	Pro	Phe	Lys	Gly
				215					220					225
Asn	Ser	Val	Ile	His	Glu	Thr	Glu	Tyr	Lys	Arg	Asn	Phe	Lys	Gly
				230					235					240
Leu	Ser	Pro	Val	Lys	Glu	Pro	Lys	Leu	Arg	Asn	Asp	Leu	Arg	Glu
				245					250					255
Asn	Arg	Asn	Leu	Glu	Thr	Val	Ser	Pro	Glu	Arg	Lys	Ser	Asn	Lys
				260					265					270
Ile	Asp	Asp	Arg	Leu	Lys	Leu	Glu	Ala	Glu	Met	Glu	Leu	Lys	Asp
				275					280					285
Leu	His	Gln	Pro	Lys	Arg	Lys	Leu	Thr	Pro	Trp	Lys	His	Gln	Arg
				290					295					300
Leu	Gly	Lys	Val	Asn	Ser	Glu	Tyr	Arg	Ala	Lys	Phe	Leu	Ser	Pro
				305					310					315
Ala	Gln	Tyr	Leu	Tyr	Lys	Ala	Gly	Ala	Trp	Thr	His	Val	Lys	Gly
				320					325					330
Asn	Met	Pro	Asn	Gln	Gly	Ser	Leu	Asn	Ala	Met	Trp	Tyr	Ala	Glu
				335					340					345
Val	Lys	Glu	Leu	Arg	Glu	Lys	Ala	Glu	Phe	Tyr	Arg	Lys	Arg	Val
				350					355					360
Gln	Gly	Thr	His	Phe	Ser	Arg	Asp	His	Leu	Asn	Gln	Ile	Leu	Ser
				365					370					375
Asp	Ser	Asn	Cys	Cys	Trp	Asp	Val	Ser	Ser	Thr	Thr	Ser	Ser	Glu



	380		385		390
Gly Thr Ile Ser	Ser Asn Ile Arg Ala	Leu Asp Leu Ala Gly Asp			
	395		400		405
Pro Thr Ser His	Lys Thr Leu Gln Lys Cys	Pro Ser Thr Glu Pro			
	410		415		420
Glu Glu Lys Gly	Asn Ile Val Glu Glu Gln	Pro Gln Lys Asn Thr			
	425		430		435
Thr Glu Lys Leu	Gly Val Ser Ala Pro Thr	Ile Pro Val Arg Arg			
	440		445		450
Arg Leu Ala Trp	Asp Thr Glu Asn Thr Ser	Glu Asp Val Gln Lys			
	455		460		465
Gln Pro Gly Glu	Lys Glu Glu Glu Asp Asp	Asn Glu Glu Glu Gly			
	470		475		480
Asp Arg Lys Thr	Gly Lys Gln Ala Phe Met	Gly Glu Gln Glu Lys			
	485		490		495
Leu Asp Val His	Glu Lys Ser Lys Ala Asp	Lys Met Lys Glu Gly			
	500		505		510
Ser Asp Ser Ser	Val Ser Ser Glu Lys Gly	Gly Arg Leu Pro Thr			
	515		520		525
Pro Lys Leu Arg	Glu Leu Gly Gly Ile Gln	Arg Thr His His Asp			
	530		535		540
Leu Thr Thr Pro	Ala Val Gly Gly Ala Val	Leu Val Ser Pro Ser			
	545		550		555
Lys Met Lys Pro	Pro Ala Leu Glu Gln Arg	Lys Arg Met Thr Ser			
	560		565		570
Gln Asp Cys Leu	Glu Thr Ser Lys Asn Asp	Phe Thr Lys Lys Glu			
	575		580		585
Ser Arg Ala Val	Ser Leu Leu Thr Ser Pro	Ala Ala Gly Ile Lys			
	590		595		600
Thr Val Asp Pro	Leu Pro Leu Arg Glu Asp	Ser Glu Asp Asn Ile			
	605		610		615
His Lys Phe Ala	Glu Ala Thr Leu Pro Val	Ser Lys Ile Pro Lys			
	620		625		630
Tyr Pro Thr Asn	Pro Pro Gly Gln Leu Pro	Ser Pro Pro His Val			
	635		640		645
Pro Ser Tyr Trp	His Pro Ser Arg Arg Ile	Gln Gly Ser Leu Arg			
	650		655		660
Asp Pro Glu Phe	Gln His Asn Val Gly Lys	Ala Arg Met Asn Asn			
	665		670		675
Leu Gln Leu Pro	Gln His Glu Ala Phe Asn	Asp Glu Asp Glu Asp			
	680		685		690
Arg Leu Ser Glu	Ile Ser Ala Arg Ser Ala	Ala Ser Ser Leu Arg			
	695		700		705
Ala Phe Gln Thr	Leu Ala Arg Ala Lys Lys	Arg Lys Glu Asn Phe			
	710		715		720
Trp Gly Lys Thr					

&lt;210&gt; 37

&lt;211&gt; 605

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 8268822CD1

&lt;400&gt; 37

```

Met Glu Ile Gly Thr Glu Ile Ser Arg Lys Ile Arg Ser Ala Ile
  1           5           10           15
Lys Gly Lys Leu Gln Glu Leu Gly Ala Tyr Val Asp Glu Glu Leu
  20           25           30
Pro Asp Tyr Ile Met Val Met Val Ala Asn Lys Lys Ser Gln Asp
  35           40           45
Gln Met Thr Glu Asp Leu Ser Leu Phe Leu Gly Asn Asn Thr Ile
  50           55           60
Arg Phe Thr Val Trp Leu His Gly Val Leu Asp Lys Leu Arg Ser
  65           70           75
Val Thr Thr Glu Pro Ser Ser Leu Lys Ser Ser Asp Thr Asn Ile
  80           85           90
Phe Asp Ser Asn Val Pro Ser Asn Lys Ser Asn Phe Ser Arg Gly
  95           100          105
Asp Glu Arg Arg His Glu Ala Ala Val Pro Pro Leu Ala Ile Pro
  110          115          120
Ser Ala Arg Pro Glu Lys Arg Asp Ser Arg Val Ser Thr Ser Ser
  125          130          135
Gln Glu Ser Lys Thr Thr Asn Val Arg Gln Thr Tyr Asp Asp Gly
  140          145          150
Ala Ala Thr Arg Leu Met Ser Thr Val Lys Pro Leu Arg Glu Pro
  155          160          165
Ala Pro Ser Glu Asp Val Ile Asp Ile Lys Pro Glu Pro Asp Asp
  170          175          180
Leu Ile Asp Glu Asp Leu Asn Phe Val Gln Glu Asn Pro Leu Ser
  185          190          195
Gln Lys Lys Pro Thr Val Thr Leu Thr Tyr Gly Ser Ser Arg Pro
  200          205          210
Ser Ile Glu Ile Tyr Arg Pro Pro Ala Ser Arg Asn Ala Asp Ser
  215          220          225
Gly Val His Leu Asn Arg Leu Gln Phe Gln Gln Gln Gln Asn Ser
  230          235          240
Ile His Ala Ala Lys Gln Leu Asp Met Gln Ser Ser Trp Val Tyr
  245          250          255
Glu Thr Gly Arg Leu Cys Glu Pro Glu Val Leu Asn Ser Leu Glu
  260          265          270
Glu Thr Tyr Ser Pro Phe Phe Arg Asn Asn Ser Glu Lys Met Ser
  275          280          285
Met Glu Asp Glu Asn Phe Arg Lys Arg Lys Leu Pro Val Val Ser
  290          295          300
Ser Val Val Lys Val Lys Lys Phe Asn His Asp Gly Glu Glu Glu
  305          310          315
Glu Glu Asp Asp Asp Tyr Gly Ser Arg Thr Gly Ser Ile Ser Ser
  320          325          330
Ser Val Ser Val Pro Ala Lys Pro Glu Arg Arg Pro Ser Leu Pro
  335          340          345
Pro Ser Lys Gln Ala Asn Lys Asn Leu Ile Leu Lys Ala Ile Ser
  350          355          360
Glu Ala Gln Glu Ser Val Thr Lys Thr Thr Asn Tyr Ser Thr Val
  365          370          375
Pro Gln Lys Gln Thr Leu Pro Val Ala Pro Arg Thr Arg Thr Ser
  380          385          390
Gln Glu Glu Leu Leu Ala Glu Val Val Gln Gly Gln Ser Arg Thr
  395          400          405
Pro Arg Ile Ser Pro Pro Ile Lys Glu Glu Glu Thr Lys Gly Asp

```

	410		415		420
Ser Val Glu Lys Asn Gln Gly Thr Gln	Gln Arg Gln Leu Leu Ser				
	425		430		435
Arg Leu Gln Ile Asp Pro Val Met Ala	Glu Thr Leu Gln Met Ser				
	440		445		450
Gln Ala Glu Met Ser Glu Leu Ser Val	Ala Gln Lys Pro Glu Lys				
	455		460		465
Leu Leu Glu Arg Cys Lys Tyr Trp Pro	Ala Cys Lys Asn Gly Asp				
	470		475		480
Glu Cys Ala Tyr His His Pro Ile Ser	Pro Cys Lys Ala Phe Pro				
	485		490		495
Asn Cys Lys Phe Ala Glu Lys Cys Leu	Phe Val His Pro Asn Cys				
	500		505		510
Lys Tyr Asp Ala Lys Cys Thr Lys Pro	Asp Cys Pro Phe Thr His				
	515		520		525
Val Ser Arg Arg Ile Pro Val Leu Ser	Pro Lys Pro Ala Val Ala				
	530		535		540
Pro Pro Ala Pro Pro Ser Ser Ser Gln	Leu Cys Arg Tyr Phe Pro				
	545		550		555
Ala Cys Lys Lys Met Glu Cys Pro Phe	Tyr His Pro Lys His Cys				
	560		565		570
Arg Phe Asn Thr Gln Cys Thr Arg Pro	Asp Cys Thr Phe Tyr His				
	575		580		585
Pro Thr Ile Asn Val Pro Pro Arg His	Ala Leu Lys Trp Ile Arg				
	590		595		600
Pro Gln Thr Ser Glu					
	605				

&lt;210&gt; 38

&lt;211&gt; 865

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1814553CD1

&lt;400&gt; 38

Met Thr Glu Val Lys Val Lys Thr Glu Leu Pro Asp Asp Tyr Ile		
1	5	10
Gln Glu Val Ile Trp Gln Gly Glu Ala Lys Glu Glu Lys Lys Ala		
20	25	30
Val Ser Lys Asp Gly Thr Ser Asp Val Pro Ala Glu Ile Cys Val		
35	40	45
Val Ile Gly Gly Val Arg Asn Gln Gln Thr Leu Gly Ser Tyr Glu		
50	55	60
Cys Gly Ile Cys Gly Lys Lys Tyr Lys Tyr Tyr Asn Cys Phe Gln		
65	70	75
Thr His Val Arg Ala His Arg Asp Thr Glu Ala Thr Ser Gly Glu		
80	85	90
Gly Ala Ser Gln Ser Asn Asn Phe Arg Tyr Thr Cys Asp Ile Cys		
95	100	105
Gly Lys Lys Tyr Lys Tyr Tyr Ser Cys Phe Gln Glu His Arg Asp		
110	115	120
Leu His Ala Val Asp Val Phe Ser Val Glu Gly Ala Pro Glu Asn		
125	130	135

Arg	Ala	Asp	Pro	Phe	Asp	Gln	Gly	Val	Val	Ala	Thr	Asp	Glu	Val	140	145	150
Lys	Glu	Glu	Pro	Pro	Glu	Pro	Phe	Gln	Lys	Ile	Gly	Pro	Lys	Thr	155	160	165
Gly	Asn	Tyr	Thr	Cys	Glu	Phe	Cys	Gly	Lys	Gln	Tyr	Lys	Tyr	Tyr	170	175	180
Thr	Pro	Tyr	Gln	Glu	His	Val	Ala	Leu	His	Ala	Pro	Ile	Ser	Thr	185	190	195
Ala	Pro	Gly	Trp	Glu	Pro	Pro	Asp	Asp	Pro	Asp	Thr	Gly	Ser	Glu	200	205	210
Cys	Ser	His	Pro	Glu	Val	Ser	Pro	Ser	Pro	Arg	Phe	Val	Ala	Ala	215	220	225
Lys	Thr	Gln	Thr	Asn	Gln	Ser	Gly	Lys	Lys	Ala	Pro	Ala	Ser	Val	230	235	240
Val	Arg	Cys	Ala	Thr	Leu	Leu	His	Arg	Thr	Pro	Pro	Ala	Thr	Gln	245	250	255
Thr	Gln	Thr	Phe	Arg	Thr	Pro	Asn	Ser	Gly	Ser	Pro	Ala	Ser	Lys	260	265	270
Ala	Thr	Ala	Glu	Ser	Ala	Phe	Ser	Arg	Arg	Val	Glu	Gly	Lys	Ala	275	280	285
Gln	Asn	His	Phe	Glu	Glu	Thr	Asn	Ser	Ser	Ser	Gln	Asn	Ser	Ser	290	295	300
Glu	Pro	Tyr	Thr	Cys	Gly	Ala	Cys	Gly	Ile	Gln	Phe	Gln	Phe	Tyr	305	310	315
Asn	Asn	Leu	Leu	Glu	His	Met	Gln	Ser	His	Ala	Ala	Asp	Asn	Glu	320	325	330
Asn	Asn	Ile	Ala	Ser	Asn	Gln	Ser	Arg	Ser	Pro	Pro	Ala	Val	Val	335	340	345
Glu	Glu	Lys	Trp	Lys	Pro	Gln	Ala	Gln	Arg	Asn	Ser	Ala	Asn	Asn	350	355	360
Thr	Thr	Thr	Ser	Gly	Leu	Thr	Pro	Asn	Ser	Met	Ile	Pro	Glu	Lys	365	370	375
Glu	Arg	Gln	Asn	Ile	Ala	Glu	Arg	Leu	Leu	Arg	Val	Met	Cys	Ala	380	385	390
Asp	Leu	Gly	Ala	Leu	Ser	Val	Val	Ser	Gly	Lys	Glu	Phe	Leu	Lys	395	400	405
Leu	Ala	Gln	Thr	Leu	Val	Asp	Ser	Gly	Ala	Arg	Tyr	Gly	Ala	Phe	410	415	420
Ser	Val	Thr	Glu	Ile	Leu	Gly	Asn	Phe	Asn	Thr	Leu	Ala	Leu	Lys	425	430	435
His	Leu	Pro	Arg	Met	Tyr	Asn	Gln	Val	Lys	Val	Lys	Val	Thr	Cys	440	445	450
Ala	Leu	Gly	Ser	Asn	Ala	Cys	Leu	Gly	Ile	Gly	Val	Thr	Cys	His	455	460	465
Ser	Gln	Ser	Val	Gly	Pro	Asp	Ser	Cys	Tyr	Ile	Leu	Thr	Ala	Tyr	470	475	480
Gln	Ala	Glu	Gly	Asn	His	Ile	Lys	Ser	Tyr	Val	Leu	Gly	Val	Lys	485	490	495
Gly	Ala	Asp	Ile	Arg	Asp	Ser	Gly	Asp	Leu	Val	His	His	Trp	Val	500	505	510
Gln	Asn	Val	Leu	Ser	Glu	Phe	Val	Met	Ser	Glu	Ile	Arg	Thr	Val	515	520	525
Tyr	Val	Thr	Asp	Cys	Arg	Val	Ser	Thr	Ser	Ala	Phe	Ser	Lys	Ala	530	535	540
Gly	Met	Cys	Leu	Arg	Cys	Ser	Ala	Cys	Ala	Leu	Asn	Ser	Val	Val	545	550	555

Gln Ser Val Leu Ser Lys Arg Thr Leu Gln Ala Arg Ser Met His	560	565	570
Glu Val Ile Glu Leu Leu Asn Val Cys Glu Asp Leu Ala Gly Ser	575	580	585
Thr Gly Leu Ala Lys Glu Thr Phe Gly Ser Leu Glu Glu Thr Ser	590	595	600
Pro Pro Pro Cys Trp Asn Ser Val Thr Asp Ser Leu Leu Leu Val	605	610	615
His Glu Arg Tyr Glu Gln Ile Cys Glu Phe Tyr Ser Arg Ala Lys	620	625	630
Lys Met Asn Leu Ile Gln Ser Leu Asn Lys His Leu Leu Ser Asn	635	640	645
Leu Ala Ala Ile Leu Thr Pro Val Lys Gln Ala Val Ile Glu Leu	650	655	660
Ser Asn Glu Ser Gln Pro Thr Leu Gln Leu Val Leu Pro Thr Tyr	665	670	675
Val Arg Leu Glu Lys Leu Phe Thr Ala Lys Ala Asn Asp Ala Gly	680	685	690
Thr Val Ser Lys Leu Cys His Leu Phe Leu Glu Ala Leu Lys Glu	695	700	705
Asn Phe Lys Val His Pro Ala His Lys Val Ala Met Ile Leu Asp	710	715	720
Pro Gln Gln Lys Leu Arg Pro Val Pro Pro Tyr Gln His Glu Glu	725	730	735
Ile Ile Gly Lys Val Cys Glu Leu Ile Asn Glu Val Lys Glu Ser	740	745	750
Trp Ala Glu Glu Ala Asp Phe Glu Pro Ala Ala Lys Lys Pro Arg	755	760	765
Ser Ala Ala Val Glu Asn Pro Ala Ala Gln Glu Asp Asp Arg Leu	770	775	780
Gly Lys Asn Glu Val Tyr Asp Tyr Leu Gln Glu Pro Leu Phe Gln	785	790	795
Ala Thr Pro Asp Leu Phe Gln Tyr Trp Ser Cys Val Thr Gln Lys	800	805	810
His Thr Lys Leu Ala Lys Leu Ala Phe Trp Leu Leu Ala Val Pro	815	820	825
Ala Val Gly Ala Arg Ser Gly Cys Val Asn Met Cys Glu Gln Ala	830	835	840
Leu Leu Ile Lys Arg Arg Arg Leu Leu Ser Pro Glu Asp Met Asn	845	850	855
Lys Leu Met Phe Leu Lys Ser Asn Met Leu	860	865	

&lt;210&gt; 39

&lt;211&gt; 219

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 71217830CD1

&lt;400&gt; 39

Met Tyr Ile Tyr Thr Phe Leu Ser His Leu Val Lys Cys Leu Lys	1	5	10	15
Ile Arg Val Ile Phe Asn Leu Ile Phe Leu Ile Leu Gly Lys Tyr				

	20		25		30
Ile	Ala	Ser	Thr	Gln	Arg
	35		40		45
Arg	Val	Lys	Glu	Gly	Tyr
	50		55		60
Glu	Asn	Lys	Tyr	Val	Lys
	65		70		75
Pro	Gly	Leu	Ser	Pro	Glu
	80		85		90
Pro	Glu	Gly	Gly	Glu	Pro
	95		100		105
Leu	Lys	Arg	Lys	Glu	Lys
	110		115		120
Ala	Glu	Ala	Leu	Ser	Arg
	125		130		135
Thr	Ala	Gln	Leu	Pro	Ser
	140		145		150
Thr	Ala	Ala	Ser	Asp	Gln
	155		160		165
Ala	Lys	Lys	Ile	Lys	Asn
	170		175		180
Glu	Leu	Gln	Gln	Arg	Ile
	185		190		195
Lys	Glu	Gln	Leu	Glu	Lys
	200		205		210
Glu	Leu	Glu	Asp	Leu	Glu
	215				

&lt;210&gt; 40

&lt;211&gt; 1144

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7506252CD1

&lt;400&gt; 40

Met	Glu	Arg	Leu	Arg	Asp	Val	Arg	Glu	Arg	Leu	Gln	Ala	Trp	Glu
1			5					10						15
Arg	Ala	Phe	Arg	Arg	Gln	Arg	Gly	Arg	Arg	Pro	Ser	Gln	Asp	Asp
			20					25						30
Val	Glu	Ala	Ala	Pro	Glu	Glu	Thr	Arg	Ala	Leu	Tyr	Arg	Glu	Tyr
			35					40						45
Arg	Thr	Leu	Lys	Arg	Thr	Thr	Gly	Gln	Ala	Gly	Gly	Gly	Leu	Arg
			50					55						60
Ser	Ser	Glu	Ser	Leu	Pro	Ala	Ala	Ala	Glu	Glu	Ala	Pro	Glu	Pro
			65					70						75
Arg	Cys	Trp	Gly	Pro	His	Leu	Asn	Arg	Ala	Ala	Thr	Lys	Ser	Pro
			80					85						90
Gln	Pro	Thr	Pro	Gly	Arg	Ser	Arg	Gln	Gly	Ser	Val	Pro	Asp	Tyr
			95					100						105
Gly	Gln	Arg	Leu	Lys	Ala	Asn	Leu	Lys	Gly	Thr	Leu	Gln	Ala	Gly
			110					115						120
Pro	Ala	Leu	Gly	Arg	Arg	Pro	Trp	Pro	Leu	Gly	Arg	Ala	Ser	Ser
			125					130						135

Lys	Ala	Ser	Thr	Pro	Lys	Pro	Pro	Gly	Thr	Gly	Pro	Val	Pro	Ser
				140					145					150
Phe	Ala	Glu	Lys	Val	Ser	Asp	Glu	Pro	Pro	Gln	Leu	Pro	Glu	Pro
				155					160					165
Gln	Pro	Arg	Pro	Gly	Arg	Leu	Gln	His	Leu	Gln	Ala	Ser	Leu	Ser
				170					175					180
Gln	Arg	Leu	Gly	Ser	Leu	Asp	Pro	Gly	Trp	Leu	Gln	Arg	Cys	His
				185					190					195
Ser	Glu	Val	Pro	Asp	Phe	Leu	Gly	Ala	Pro	Lys	Ala	Cys	Arg	Pro
				200					205					210
Asp	Leu	Gly	Ser	Glu	Glu	Ser	Gln	Leu	Leu	Ile	Pro	Gly	Glu	Ser
				215					220					225
Ala	Val	Leu	Gly	Pro	Gly	Ala	Gly	Ser	Gln	Gly	Pro	Glu	Ala	Ser
				230					235					240
Ala	Phe	Gln	Glu	Val	Ser	Ile	Arg	Val	Gly	Ser	Pro	Gln	Pro	Ser
				245					250					255
Ser	Ser	Gly	Gly	Glu	Lys	Arg	Arg	Trp	Asn	Glu	Glu	Pro	Trp	Glu
				260					265					270
Ser	Pro	Ala	Gln	Val	Gln	Gln	Glu	Ser	Ser	Gln	Ala	Gly	Pro	Pro
				275					280					285
Ser	Glu	Gly	Ala	Gly	Ala	Val	Ala	Val	Glu	Glu	Asp	Pro	Pro	Gly
				290					295					300
Glu	Pro	Val	Gln	Ala	Gln	Pro	Pro	Gln	Pro	Cys	Ser	Ser	Pro	Ser
				305					310					315
Asn	Pro	Arg	Tyr	His	Gly	Leu	Ser	Pro	Ser	Ser	Gln	Ala	Arg	Ala
				320					325					330
Gly	Lys	Ala	Glu	Gly	Thr	Ala	Pro	Leu	His	Ile	Phe	Pro	Arg	Leu
				335					340					345
Ala	Arg	His	Asp	Arg	Gly	Asn	Tyr	Val	Arg	Leu	Asn	Met	Lys	Gln
				350					355					360
Lys	His	Tyr	Val	Arg	Gly	Arg	Ala	Leu	Arg	Ser	Arg	Leu	Leu	Arg
				365					370					375
Lys	Gln	Ala	Trp	Lys	Gln	Lys	Trp	Arg	Lys	Lys	Gly	Glu	Cys	Phe
				380					385					390
Gly	Gly	Gly	Gly	Ala	Thr	Val	Thr	Thr	Lys	Glu	Ser	Cys	Phe	Leu
				395					400					405
Asn	Glu	Gln	Phe	Asp	His	Trp	Ala	Ala	Gln	Cys	Pro	Arg	Pro	Ala
				410					415					420
Ser	Glu	Glu	Asp	Thr	Asp	Ala	Val	Gly	Pro	Glu	Pro	Leu	Val	Pro
				425					430					435
Ser	Pro	Gln	Pro	Val	Pro	Glu	Val	Pro	Ser	Leu	Asp	Pro	Thr	Val
				440					445					450
Leu	Pro	Leu	Tyr	Ser	Leu	Gly	Pro	Ser	Gly	Gln	Leu	Ala	Glu	Thr
				455					460					465
Pro	Ala	Glu	Val	Phe	Gln	Ala	Leu	Glu	Gln	Leu	Gly	His	Gln	Ala
				470					475					480
Phe	Arg	Pro	Gly	Gln	Glu	Arg	Ala	Val	Met	Arg	Ile	Leu	Ser	Gly
				485					490					495
Ile	Ser	Thr	Leu	Leu	Val	Leu	Pro	Thr	Gly	Ala	Gly	Lys	Ser	Leu
				500					505					510
Cys	Tyr	Gln	Leu	Pro	Ala	Leu	Leu	Tyr	Ser	Arg	Arg	Ser	Pro	Cys
				515					520					525
Leu	Thr	Leu	Val	Val	Ser	Pro	Leu	Leu	Ser	Leu	Met	Asp	Asp	Gln
				530					535					540
Val	Ser	Gly	Leu	Pro	Pro	Cys	Leu	Lys	Ala	Ala	Cys	Ile	His	Ser
				545					550					555

Gly Met Thr Arg	Lys Gln Arg Glu Ser Val	Leu Gln Lys Ile Arg
560	565	570
Ala Ala Gln Val	His Val Leu Met Leu Thr	Pro Glu Ala Leu Val
575	580	585
Gly Ala Gly Gly	Leu Pro Pro Ala Ala	Gln Leu Pro Pro Val Ala
590	595	600
Phe Ala Cys Ile	Asp Glu Ala His Cys	Leu Ser Gln Trp Ser His
605	610	615
Asn Phe Arg Pro	Cys Tyr Leu Arg Val	Cys Lys Val Leu Arg Glu
620	625	630
Arg Met Gly Val	His Cys Phe Leu Gly	Leu Thr Ala Thr Ala Thr
635	640	645
Arg Arg Thr Ala	Ser Asp Val Ala Gln	His Leu Ala Val Ala Glu
650	655	660
Glu Pro Asp Leu	His Gly Pro Ala Pro	Val Pro Thr Asn Leu His
665	670	675
Leu Ser Val Ser	Met Asp Arg Asp Thr	Asp Gln Ala Leu Leu Thr
680	685	690
Leu Leu Gln Gly	Lys Arg Phe Gln Asn	Leu Asp Ser Ile Ile Ile
695	700	705
Tyr Cys Asn Arg	Arg Glu Asp Thr Glu	Arg Ile Ala Ala Leu Leu
710	715	720
Arg Thr Cys Leu	His Ala Ala Trp Val	Pro Gly Ser Gly Gly Arg
725	730	735
Ala Pro Lys Thr	Thr Ala Glu Ala Tyr	His Ala Gly Met Cys Ser
740	745	750
Arg Glu Arg Arg	Arg Pro Gln Gly Glu	Asp Leu Arg Glu Leu Arg
755	760	765
Arg His Val His	Ala Asp Ser Thr Asp	Phe Leu Ala Val Lys Arg
770	775	780
Leu Val Gln Arg	Val Phe Pro Ala Cys	Thr Cys Thr Cys Thr Arg
785	790	795
Pro Pro Ser Glu	Gln Glu Gly Ala Val	Gly Gly Glu Arg Pro Val
800	805	810
Pro Lys Tyr Pro	Pro Gln Glu Ala Glu	Gln Leu Ser His Gln Ala
815	820	825
Ala Pro Gly Pro	Arg Arg Val Cys Met	Gly His Glu Arg Ala Leu
830	835	840
Pro Ile Gln Leu	Thr Val Gln Ala Leu	Asp Met Pro Glu Glu Ala
845	850	855
Ile Glu Thr Leu	Leu Cys Tyr Leu Glu	Leu His Pro His His Trp
860	865	870
Leu Glu Leu Leu	Ala Thr Thr Tyr Thr	His Cys Arg Leu Asn Cys
875	880	885
Pro Gly Gly Pro	Ala Gln Leu Gln Ala	Leu Ala His Arg Cys Pro
890	895	900
Pro Leu Ala Val	Cys Leu Ala Gln Gln	Leu Pro Glu Asp Pro Gly
905	910	915
Gln Gly Ser Ser	Ser Val Glu Phe Asp	Met Val Lys Leu Val Asp
920	925	930
Ser Met Gly Trp	Glu Leu Ala Ser Val	Arg Arg Ala Leu Cys Gln
935	940	945
Leu Gln Trp Asp	His Glu Pro Arg Thr	Gly Val Arg Arg Gly Thr
950	955	960
Gly Val Leu Val	Glu Phe Ser Glu Leu	Ala Phe His Leu Arg Ser
965	970	975



```

Pro Gly Asp Leu Thr Ala Glu Glu Lys Asp Gln Ile Cys Asp Phe
      980                      985                      990
Leu Tyr Gly Arg Val Gln Ala Arg Glu Arg Gln Ala Leu Ala Arg
      995                      1000                     1005
Leu Arg Arg Thr Phe Gln Ala Phe His Ser Val Ala Phe Pro Ser
      1010                     1015                     1020
Cys Gly Pro Cys Leu Glu Gln Gln Asp Glu Glu Arg Ser Thr Arg
      1025                     1030                     1035
Leu Lys Asp Leu Leu Gly Arg Tyr Phe Glu Glu Glu Glu Gly Gln
      1040                     1045                     1050
Glu Pro Gly Gly Met Glu Asp Ala Gln Gly Pro Glu Pro Gly Gln
      1055                     1060                     1065
Ala Arg Leu Gln Asp Trp Glu Asp Gln Val Arg Cys Asp Ile Arg
      1070                     1075                     1080
Gln Phe Leu Ser Leu Arg Pro Glu Glu Lys Phe Ser Ser Arg Ala
      1085                     1090                     1095
Val Ala Arg Ile Phe His Gly Ile Gly Ser Pro Cys Tyr Pro Ala
      1100                     1105                     1110
Gln Val Tyr Gly Gln Asp Arg Arg Phe Trp Arg Lys Tyr Leu His
      1115                     1120                     1125
Leu Ser Phe His Ala Leu Val Gly Leu Ala Thr Glu Glu Leu Leu
      1130                     1135                     1140
Gln Val Ala Arg

```

&lt;210&gt; 41

&lt;211&gt; 217

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2270608CD1

&lt;400&gt; 41

```

Met Ile Asn His Leu Ser Pro His Gln Ala Ala Ala Pro Val Asp
  1           5           10           15
Gln Thr Pro Arg Thr Leu Ala Thr Met Gly Gln Arg Ala Leu Pro
  20          25          30
Ser Ser Leu Ala Leu Leu Ser Arg Pro Leu Ser Pro Pro Pro Ala
  35          40          45
Ala Cys Ser Gly Asp Pro Gly Cys Gly Ser Gly Ala Gly Leu Pro
  50          55          60
Ser Ala Ser Ala Ala Ala Gly Ile Ala Ser Ser Ala Val Glu Pro
  65          70          75
Val Cys Gly Asp Ala Ala Pro Ala Cys Leu Leu Arg Thr Pro Leu
  80          85          90
Arg Gly Leu Leu Lys Pro Thr Gly Pro Arg Ser Thr Met Glu Cys
  95         100         105
Pro Pro Ala Leu Ile Val His Pro Pro Ala Gly Gly Met Ala Ser
  110        115        120
Gly Ser Ser Gln Pro Trp Ala Ala Ala Ser Ala Thr Pro Met Leu
  125        130        135
Ser Ser Lys Ala Ser Leu Cys Ile Pro Thr Arg Gly Pro Pro Pro
  140        145        150
Gln Pro Leu Met Arg Thr Pro Ala Ala Arg Ser His Trp Pro Ile

```

	155		160		165
Pro His Pro Cys	Asp Thr Ala Cys Pro	Ala Pro Leu Pro Val	Val		
	170		175		180
Leu Val Ala Pro	Arg Ser Thr Ile Leu	Ser Met Ser Arg Thr	Trp		
	185		190		195
Thr Cys Arg Arg	Trp Ala Val Ala Pro	Cys Arg Ala Glu Lys	Leu		
	200		205		210
Met Cys Ser Ser	Ser Arg Ser				
	215				

&lt;210&gt; 42

&lt;211&gt; 309

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7502428CD1

&lt;400&gt; 42

Met Gln Ser Leu Trp	Glu Lys Ala Cys Glu	Asn Leu Arg Asn Leu	
1	5	10	15
Asn Met Glu Thr Thr	Arg Thr Arg Cys Trp	Lys Asp Tyr Val Ser	
	20	25	30
Leu Arg Ile Glu Ala	Ile Arg Ala Glu Tyr	Gln Lys Met Pro Ala	
	35	40	45
Phe Leu His Glu Glu	Glu Gln His His Leu	Glu Arg Leu Arg Lys	
	50	55	60
Glu Gly Glu Asp Ile	Phe Gln Gln Leu Asn	Glu Ser Lys Ala Arg	
	65	70	75
Met Glu His Ser Arg	Glu Leu Leu Arg Gly	Met Tyr Glu Asp Leu	
	80	85	90
Lys Gln Met Cys His	Lys Ala Asp Val Glu	Leu Leu Gln Ala Phe	
	95	100	105
Gly Asp Ile Leu His	Arg Tyr Glu Ser Leu	Leu Leu Gln Val Ser	
	110	115	120
Glu Pro Val Asn Pro	Glu Leu Ser Ala Gly	Pro Ile Thr Gly Leu	
	125	130	135
Leu Asp Ser Leu Ser	Gly Phe Arg Val Asp	Phe Thr Leu Gln Pro	
	140	145	150
Glu Arg Ala Asn Ser	His Ile Phe Leu Cys	Gly Asp Leu Arg Ser	
	155	160	165
Met Asn Val Gly Cys	Asp Pro Gln Asp Asp	Pro Asp Ile Thr Gly	
	170	175	180
Lys Ser Glu Cys Phe	Leu Val Trp Gly Ala	Gln Ala Phe Thr Ser	
	185	190	195
Gly Lys Tyr Tyr Trp	Glu Val His Met Gly	Asp Ser Trp Asn Trp	
	200	205	210
Ala Phe Gly Val Cys	Asn Asn Tyr Trp Lys	Glu Lys Arg Gln Asn	
	215	220	225
Asp Lys Ile Asp Gly	Glu Glu Gly Leu Phe	Leu Leu Gly Cys Val	
	230	235	240
Lys Glu Asp Thr His	Cys Ser Leu Phe Thr	Thr Thr Thr Pro Leu Val	
	245	250	255
Val Gln Tyr Val Pro	Arg Pro Thr Ser Thr	Val Gly Leu Phe Leu	
	260	265	270

Asp Cys Glu Gly Arg Ile Met Ser Phe Val Asp Val Asp Gln Ser  
 275 280 285  
 Phe Leu Ile Tyr Ile Ile Pro Asn Cys Ser Phe Ser Pro Pro Leu  
 290 295 300  
 Arg Pro Ile Phe Cys Cys Ser His Phe  
 305

<210> 43

<211> 483

<212> PRT

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 368741CD1

<400> 43

Met Trp Leu Gly Thr Ser Gly Lys Ser Gly Leu Pro Gly His Cys  
 1 5 10 15  
 Leu Glu Asn Pro Leu Gln Glu Cys His Pro Ala Gln Leu Glu Glu  
 20 25 30  
 Trp Ala Leu Lys Gly Ile Ser Arg Pro Ser Val Ile Ser Gln Pro  
 35 40 45  
 Glu Gln Lys Glu Glu Pro Trp Val Leu Pro Leu Gln Asn Phe Glu  
 50 55 60  
 Ala Arg Lys Ile Pro Arg Glu Ser His Thr Asp Cys Glu His Gln  
 65 70 75  
 Val Ala Lys Leu Asn Gln Asp Asn Ser Glu Thr Ala Glu Gln Cys  
 80 85 90  
 Gly Thr Ser Ser Glu Arg Thr Asn Lys Asp Leu Ser His Thr Leu  
 95 100 105  
 Ser Trp Gly Gly Asn Trp Glu Gln Gly Leu Glu Leu Glu Gly Gln  
 110 115 120  
 Tyr Gly Thr Leu Pro Gly Glu Gly Gln Leu Glu Ser Phe Ser Gln  
 125 130 135  
 Glu Arg Asp Leu Asn Lys Leu Leu Asp Gly Tyr Val Gly Glu Lys  
 140 145 150  
 Pro Met Cys Ala Glu Cys Gly Lys Ser Phe Asn Gln Ser Ser Tyr  
 155 160 165  
 Leu Ile Arg His Leu Arg Thr His Thr Gly Glu Arg Pro Tyr Thr  
 170 175 180  
 Cys Ile Glu Cys Gly Lys Gly Phe Lys Gln Ser Ser Asp Leu Val  
 185 190 195  
 Thr His Arg Arg Thr His Thr Gly Glu Lys Pro Tyr Gln Cys Lys  
 200 205 210  
 Gly Cys Glu Lys Lys Phe Ser Asp Ser Ser Thr Leu Ile Lys His  
 215 220 225  
 Gln Arg Thr His Thr Gly Glu Arg Pro Tyr Glu Cys Pro Glu Cys  
 230 235 240  
 Gly Lys Thr Phe Gly Arg Lys Pro His Leu Ile Met His Gln Arg  
 245 250 255  
 Thr His Thr Gly Glu Lys Pro Tyr Ala Cys Leu Glu Cys His Lys  
 260 265 270  
 Ser Phe Ser Arg Ser Ser Asn Phe Ile Thr His Gln Arg Thr His  
 275 280 285  
 Thr Gly Val Lys Pro Tyr Arg Cys Asn Asp Cys Gly Glu Ser Phe

	290		295		300									
Ser	Gln	Ser	Ser	Asp	Leu	Ile	Lys	His	Gln	Arg	Thr	His	Thr	Gly
	305								310					315
Glu	Arg	Pro	Phe	Lys	Cys	Pro	Glu	Cys	Gly	Lys	Gly	Phe	Arg	Asp
	320								325					330
Ser	Ser	His	Phe	Val	Ala	His	Met	Ser	Thr	His	Ser	Gly	Glu	Arg
	335								340					345
Pro	Phe	Ser	Cys	Pro	Asp	Cys	His	Lys	Ser	Phe	Ser	Gln	Ser	Ser
	350								355					360
His	Leu	Val	Thr	His	Gln	Arg	Thr	His	Thr	Gly	Glu	Arg	Pro	Phe
	365								370					375
Lys	Cys	Glu	Asn	Cys	Gly	Lys	Gly	Phe	Ala	Asp	Ser	Ser	Ala	Leu
	380								385					390
Ile	Lys	His	Gln	Arg	Ile	His	Thr	Gly	Glu	Arg	Pro	Tyr	Lys	Cys
	395								400					405
Gly	Glu	Cys	Gly	Lys	Ser	Phe	Asn	Gln	Ser	Ser	His	Phe	Ile	Thr
	410								415					420
His	Gln	Arg	Ile	His	Leu	Gly	Asp	Arg	Pro	Tyr	Arg	Cys	Pro	Glu
	425								430					435
Cys	Gly	Lys	Thr	Phe	Asn	Gln	Arg	Ser	His	Phe	Leu	Thr	His	Gln
	440								445					450
Arg	Thr	His	Thr	Gly	Glu	Lys	Pro	Phe	His	Cys	Ser	Lys	Cys	Asn
	455								460					465
Lys	Ser	Phe	Arg	Gln	Lys	Ala	His	Leu	Leu	Cys	His	Gln	Asn	Thr
	470								475					480
His	Leu	Ile												

&lt;210&gt; 44

&lt;211&gt; 137

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7506379CD1

&lt;400&gt; 44

Met	Lys	Ala	Ser	Gly	Thr	Leu	Arg	Glu	Tyr	Lys	Leu	Lys	Lys	Met
1				5					10					15
Lys	Lys	Ser	Ser	Gly	Glu	Ile	Val	Tyr	Cys	Gly	Gln	Val	Phe	Glu
				20					25					30
Lys	Ser	Pro	Leu	Arg	Val	Lys	Asn	Phe	Gly	Ile	Trp	Leu	Arg	Tyr
				35					40					45
Asp	Ser	Arg	Ser	Gly	Thr	His	Asn	Met	Tyr	Arg	Glu	Tyr	Arg	Asp
				50					55					60
Leu	Thr	Thr	Ala	Gly	Ala	Val	Thr	Gln	Cys	Tyr	Arg	Asp	Met	Gly
				65					70					75
Ala	Arg	His	Arg	Ala	Arg	Ala	His	Ser	Ile	Gln	Ile	Met	Lys	Val
				80					85					90
Glu	Glu	Ile	Ala	Ala	Ser	Lys	Cys	Arg	Arg	Pro	Ala	Val	Lys	Gln
				95					100					105
Phe	His	Asp	Ser	Lys	Ile	Lys	Phe	Pro	Leu	Pro	His	Arg	Val	Leu
				110					115					120
Arg	Arg	Gln	His	Lys	Pro	Arg	Phe	Thr	Thr	Lys	Arg	Pro	Asn	Thr
				125					130					135

Phe Phe

&lt;210&gt; 45

&lt;211&gt; 200

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7506253CD1

&lt;400&gt; 45

```

Met Asn Asn Lys Phe Asp Ala Leu Lys Asp Asp Asp Ser Gly Asp
 1          5          10          15
His Asp Gln Asn Glu Glu Asn Ser Thr Gln Lys Asp Gly Glu Lys
          20          25          30
Glu Lys Thr Glu Arg Asp Lys Asn Gln Ser Ser Ser Lys Arg Lys
          35          40          45
Val Glu Gln Phe Trp Arg Phe Tyr Ser His Met Val Arg Pro Gly
          50          55          60
Asp Leu Thr Gly His Ser Asp Phe His Leu Phe Lys Glu Gly Ile
          65          70          75
Lys Pro Met Trp Glu Asp Asp Ala Asn Lys Asn Gly Gly Lys Trp
          80          85          90
Ile Ile Arg Leu Arg Lys Gly Leu Ala Ser Arg Cys Trp Glu Asn
          95          100         105
Leu Ile Leu Ala Met Leu Gly Glu Gln Phe Met Val Gly Glu Glu
          110         115         120
Ile Cys Gly Ala Val Ser Val Arg Phe Gln Glu Asp Ile Ile
          125         130         135
Ser Ile Trp Asn Lys Thr Ala Ser Asp Gln Ala Thr Thr Ala Arg
          140         145         150
Ile Arg Asp Thr Leu Arg Arg Val Leu Asn Leu Pro Pro Asn Thr
          155         160         165
Ile Met Glu Tyr Lys Thr His Thr Asp Ser Ile Lys Met Pro Gly
          170         175         180
Arg Leu Gly Pro Gln Arg Leu Leu Phe Gln Asn Leu Trp Lys Pro
          185         190         195
Arg Leu Asn Val Pro
          200

```

&lt;210&gt; 46

&lt;211&gt; 123

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7506353CD1

&lt;400&gt; 46

```

Met Glu Phe Pro Asp Leu Gly Lys His Cys Ser Glu Lys Thr Cys
 1          5          10          15
Lys Gln Leu Asp Phe Leu Pro Val Lys Cys Asp Ala Cys Lys Gln
          20          25          30

```

```

Asp Phe Cys Lys Asp His Phe Pro Tyr Ala Ala His Lys Cys Pro
      35                      40                      45
Phe Ala Phe Gln Lys Asp Val His Val Pro Val Cys Pro Leu Cys
      50                      55                      60
Asn Thr Pro Ile Pro Val Lys Lys Gly Gln Ile Pro Asp Val Val
      65                      70                      75
Val Gly Asp His Ile Asp Arg Asp Cys Asp Ser His Pro Gly Lys
      80                      85                      90
Lys Lys Glu Lys His Arg His Pro Leu Asp His Ser Cys Arg His
      95                      100                     105
Gly Ser Arg Pro Thr Ile Lys Ala Gly Val Ser Ser Ser Val Val
      110                     115                     120
Gly Ser Gly

```

&lt;210&gt; 47

&lt;211&gt; 874

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7506372CD1

&lt;400&gt; 47

```

Met Tyr Gly Ala Ser Gly Gly Arg Ala Lys Pro Glu Arg Lys Ser
  1          5          10          15
Gly Ala Lys Glu Glu Ala Gly Pro Gly Gly Ala Gly Gly Gly Gly
      20          25          30
Ser Arg Val Glu Leu Leu Val Phe Gly Tyr Ala Cys Lys Leu Phe
      35          40          45
Arg Asp Asp Glu Arg Ala Leu Ala Gln Glu Gln Gly Gln His Leu
      50          55          60
Ile Pro Trp Met Gly Asp His Lys Ile Leu Ile Asp Arg Tyr Asp
      65          70          75
Gly Arg Gly His Leu His Asp Leu Ser Glu Tyr Asp Ala Glu Tyr
      80          85          90
Ser Thr Trp Asn Arg Asp Tyr Gln Leu Ser Glu Glu Glu Ala Arg
      95          100         105
Ile Glu Ala Leu Cys Asp Glu Glu Arg Tyr Leu Ala Leu His Thr
      110         115         120
Asp Leu Leu Glu Glu Glu Ala Arg Gln Glu Glu Glu Tyr Lys Arg
      125         130         135
Leu Ser Glu Ala Leu Ala Glu Asp Gly Ser Tyr Asn Ala Val Gly
      140         145         150
Phe Thr Tyr Gly Ser Asp Tyr Tyr Asp Pro Ser Glu Pro Thr Glu
      155         160         165
Glu Glu Glu Pro Ser Lys Gln Arg Glu Lys Asn Glu Ala Glu Asn
      170         175         180
Leu Glu Glu Asn Glu Glu Pro Phe Val Ala Pro Leu Gly Leu Ser
      185         190         195
Val Pro Ser Asp Val Glu Leu Pro Pro Thr Ala Lys Met His Ala
      200         205         210
Ile Ile Glu Arg Thr Ala Ser Phe Val Cys Arg Gln Gly Ala Gln
      215         220         225
Phe Glu Ile Met Leu Lys Ala Lys Gln Ala Arg Asn Ser Gln Phe

```

	230		235		240
Asp Phe Leu Arg	Phe Asp His Tyr Leu Asn Pro Tyr Tyr Lys Phe				
	245		250		255
Ile Gln Lys Ala	Met Lys Glu Gly Arg Tyr Thr Val Leu Ala Glu				
	260		265		270
Asn Lys Ser Asp	Glu Lys Lys Lys Ser Gly Val Ser Ser Asp Asn				
	275		280		285
Glu Asp Asp Asp	Asp Glu Glu Asp Gly Asn Tyr Leu His Pro Ser				
	290		295		300
Leu Phe Ala Ser	Lys Lys Cys Asn Arg Leu Glu Glu Leu Met Lys				
	305		310		315
Pro Leu Lys Val	Val Asp Pro Asp His Pro Leu Ala Ala Leu Val				
	320		325		330
Arg Lys Ala Gln	Ala Asp Ser Ser Thr Pro Thr Pro His Asn Ala				
	335		340		345
Asp Gly Ala Pro	Val Gln Pro Ser Gln Val Glu Tyr Thr Ala Asp				
	350		355		360
Ser Thr Val Ala	Ala Met Tyr Tyr Ser Tyr Tyr Met Leu Pro Asp				
	365		370		375
Gly Thr Tyr Cys	Leu Ala Pro Pro Pro Pro Gly Ile Asp Val Thr				
	380		385		390
Thr Tyr Tyr Ser	Thr Leu Pro Ala Gly Val Thr Val Ser Asn Ser				
	395		400		405
Pro Gly Val Thr	Thr Thr Ala Pro Pro Pro Pro Gly Thr Thr Pro				
	410		415		420
Pro Pro Pro Pro	Thr Thr Ala Glu Thr Ser Ser Gly Ala Thr Ser				
	425		430		435
Thr Thr Thr Thr	Thr Ser Ala Leu Ala Pro Val Ala Ala Ile Ile				
	440		445		450
Pro Pro Pro Pro	Asp Val Gln Pro Val Ile Asp Lys Leu Ala Glu				
	455		460		465
Tyr Val Ala Arg	Asn Gly Leu Lys Phe Glu Thr Ser Val Arg Ala				
	470		475		480
Lys Asn Asp Gln	Arg Phe Glu Phe Leu Gln Pro Trp His Gln Tyr				
	485		490		495
Asn Ala Tyr Tyr	Glu Phe Lys Lys Gln Phe Phe Leu Gln Lys Glu				
	500		505		510
Gly Gly Asp Ser	Met Gln Ala Val Ser Ala Pro Glu Glu Ala Pro				
	515		520		525
Thr Asp Ser Ala	Pro Glu Lys Pro Ser Asp Ala Gly Glu Asp Gly				
	530		535		540
Ala Pro Glu Asp	Ala Ala Glu Val Gly Ala Arg Ala Gly Ser Gly				
	545		550		555
Gly Lys Lys Glu	Ala Ser Ser Ser Lys Thr Val Pro Asp Gly Lys				
	560		565		570
Leu Val Lys Ala	Lys Gln Lys Leu Glu Asp Arg Leu Ala Ala Ala				
	575		580		585
Ala Arg Glu Lys	Leu Ala Gln Ala Ser Lys Glu Ser Lys Glu Lys				
	590		595		600
Gln Leu Gln Ala	Glu Arg Lys Arg Lys Ala Ala Leu Phe Leu Gln				
	605		610		615
Thr Leu Lys Asn	Pro Leu Pro Glu Ala Glu Ala Gly Lys Ile Glu				
	620		625		630
Glu Ser Pro Phe	Ser Val Glu Glu Ser Ser Thr Thr Pro Cys Pro				
	635		640		645
Leu Leu Thr Gly	Gly Arg Pro Leu Pro Thr Leu Glu Val Lys Pro				

	650		655		660
Pro Asp Arg Pro	Ser Ser Lys Ser Lys	Asp Pro Pro Arg Glu	Glu		
	665		670		675
Glu Lys Glu Lys	Lys Lys Lys Lys His	Lys Lys Arg Ser Arg	Thr		
	680		685		690
Arg Ser Arg Ser	Pro Lys Tyr His Ser	Ser Ser Lys Ser Arg	Ser		
	695		700		705
Arg Ser His Ser	Lys Ala Lys His Ser	Leu Pro Ser Ala Tyr	Arg		
	710		715		720
Thr Val Arg Arg	Ser Arg Ser Arg Ser	Arg Ser Pro Arg Arg	Arg		
	725		730		735
Ala His Ser Pro	Glu Arg Arg Arg Glu	Glu Arg Ser Val Pro	Thr		
	740		745		750
Ala Tyr Arg Val	Ser Arg Ser Pro Gly	Ala Ser Arg Lys Arg	Thr		
	755		760		765
Arg Ser Arg Ser	Pro His Glu Lys Lys	Lys Lys Arg Arg Ser	Arg		
	770		775		780
Ser Arg Thr Lys	Ser Lys Ala Arg Ser	Gln Ser Val Ser Pro	Ser		
	785		790		795
Lys Gln Ala Ala	Pro Arg Pro Ala Ala	Pro Ala Ala His Ser	Ala		
	800		805		810
His Ser Ala Ser	Val Ser Pro Val Glu	Ser Arg Gly Ser Ser	Gln		
	815		820		825
Glu Arg Ser Arg	Gly Val Ser Gln Glu	Lys Glu Ala Gln Ile	Ser		
	830		835		840
Ser Ala Ile Val	Ser Ser Val Gln Ser	Lys Ile Thr Gln Asp	Leu		
	845		850		855
Met Ala Lys Val	Arg Ala Met Leu Ala	Ala Ser Lys Asn Leu	Gln		
	860		865		870
Thr Ser Ala Ser					

&lt;210&gt; 48

&lt;211&gt; 189

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7506335CD1

&lt;400&gt; 48

Met Ala Ala Asp Thr	Pro Gly Lys Pro Ser	Ala Ser Pro Met	Ala
1	5	10	15
Gly Ala Pro Ala Ser	Ala Ser Arg Thr Pro	Asp Lys Pro Arg	Ser
	20	25	30
Ala Ala Glu His Arg	Lys Ser Ser Arg His	Ser Lys Leu Glu	Lys
	35	40	45
Ala Asp Ile Leu Glu	Met Thr Val Arg His	Leu Arg Ser Leu	Arg
	50	55	60
Arg Val Gln Val Thr	Ala Ala Leu Ser Ala	Asp Pro Ala Val	Leu
	65	70	75
Gly Lys Tyr Arg Ala	Gly Phe His Glu Cys	Leu Ala Glu Val	Asn
	80	85	90
Arg Phe Leu Ala Gly	Cys Glu Gly Val Pro	Ala Asp Val Arg	Ser
	95	100	105



```

Arg Leu Leu Gly His Leu Ala Ala Cys Leu Arg Gln Leu Gly Pro
      110              115              120
Ser Arg Arg Pro Ala Ser Leu Ser Pro Ala Ala Pro Ala Glu Ala
      125              130              135
Pro Ala Pro Glu Val Tyr Ala Gly Arg Pro Leu Leu Pro Ser Leu
      140              145              150
Gly Gly Pro Phe Pro Leu Leu Ala Pro Pro Leu Leu Pro Gly Leu
      155              160              165
Thr Arg Ala Leu Pro Ala Ala Pro Arg Ala Gly Pro Gln Gly Pro
      170              175              180
Gly Gly Pro Trp Arg Pro Trp Leu Arg
      185

```

&lt;210&gt; 49

&lt;211&gt; 716

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 5546982CD1

&lt;400&gt; 49

```

Met Glu Gly Asp Gly Ser Asp Ser Pro Val Thr Ile Lys Asn Ile
  1          5          10          15
Glu Arg Glu Leu Ile Cys Pro Ala Cys Lys Glu Leu Phe Thr His
  20          25          30
Pro Leu Ile Leu Pro Cys Gln His Ser Ile Cys His Lys Cys Val
  35          40          45
Lys Glu Leu Leu Leu Thr Leu Asp Asp Ser Phe Asn Asp Val Gly
  50          55          60
Ser Asp Asn Ser Asn Gln Ser Ser Pro Arg Leu Arg Leu Pro Ser
  65          70          75
Pro Ser Met Asp Lys Ile Asp Arg Ile Asn Arg Pro Gly Trp Lys
  80          85          90
Arg Asn Ser Leu Thr Pro Arg Thr Thr Val Phe Pro Cys Pro Gly
  95          100         105
Cys Glu His Asp Val Asp Leu Gly Glu Arg Gly Ile Asn Gly Leu
  110         115         120
Phe Arg Asn Phe Thr Leu Glu Thr Ile Val Glu Arg Tyr Arg Gln
  125         130         135
Ala Ala Arg Ala Ala Thr Ala Ile Met Cys Asp Leu Cys Lys Pro
  140         145         150
Pro Pro Gln Glu Ser Thr Lys Ser Cys Met Asp Cys Ser Ala Ser
  155         160         165
Tyr Cys Asn Glu Cys Phe Lys Ile His His Pro Trp Gly Thr Ile
  170         175         180
Lys Ala Gln His Glu Tyr Val Gly Pro Thr Thr Asn Phe Arg Pro
  185         190         195
Lys Ile Leu Met Cys Pro Glu His Glu Thr Glu Arg Ile Asn Met
  200         205         210
Tyr Cys Glu Leu Cys Arg Arg Pro Val Cys His Leu Cys Lys Leu
  215         220         225
Gly Gly Asn His Ala Asn His Arg Val Thr Thr Met Ser Ser Ala
  230         235         240
Tyr Lys Thr Leu Lys Glu Lys Leu Ser Lys Asp Ile Asp Tyr Leu

```

	245		250		255
Ile Gly Lys Glu	Ser Gln Val Lys Ser	Gln Ile Ser Glu Leu	Asn		
	260		265		270
Leu Leu Met Lys	Glu Thr Glu Cys Asn	Gly Glu Arg Ala Lys	Glu		
	275		280		285
Glu Ala Ile Thr	His Phe Glu Lys Leu	Phe Glu Val Leu Glu	Glu		
	290		295		300
Arg Lys Ser Ser	Val Leu Lys Ala Ile	Asp Ser Ser Lys Lys	Leu		
	305		310		315
Arg Leu Asp Lys	Phe Gln Thr Gln Met	Glu Glu Tyr Gln Gly	Leu		
	320		325		330
Leu Glu Asn Asn	Gly Leu Val Gly Tyr	Ala Gln Glu Val Leu	Lys		
	335		340		345
Glu Thr Asp Gln	Ser Cys Phe Val Gln	Thr Ala Lys Gln Leu	His		
	350		355		360
Leu Arg Ile Gln	Lys Ala Thr Glu Ser	Leu Arg Ser Phe Arg	Pro		
	365		370		375
Ala Ala Gln Thr	Ser Phe Glu Asp Tyr	Val Val Asn Thr Ser	Lys		
	380		385		390
Gln Thr Glu Leu	Leu Gly Glu Leu Ser	Phe Phe Ser Ser Gly	Ile		
	395		400		405
Asp Val Pro Glu	Ile Asn Glu Glu Gln	Ser Lys Val Tyr Asn	Asn		
	410		415		420
Ala Leu Ile Asn	Trp His His Pro Glu	Lys Asp Lys Ala Asp	Ser		
	425		430		435
Tyr Val Leu Glu	Tyr Arg Lys Ile Asn	Arg Asp Asp Glu Met	Ser		
	440		445		450
Trp Asn Glu Ile	Glu Val Cys Gly Thr	Ser Lys Ile Ile Gln	Asp		
	455		460		465
Leu Glu Asn Ser	Ser Thr Tyr Ala Phe	Arg Val Arg Ala Tyr	Lys		
	470		475		480
Gly Ser Ile Cys	Ser Pro Cys Ser Arg	Glu Leu Ile Leu His	Thr		
	485		490		495
Pro Pro Ala Pro	Val Phe Ser Phe Leu	Phe Asp Glu Lys Cys	Gly		
	500		505		510
Tyr Asn Asn Glu	His Leu Leu Leu Asn	Leu Lys Arg Asp Arg	Val		
	515		520		525
Glu Ser Arg Ala	Gly Phe Asn Leu Leu	Leu Ala Ala Glu Arg	Ile		
	530		535		540
Gln Val Gly Tyr	Tyr Thr Ser Leu Asp	Tyr Ile Ile Gly Asp	Thr		
	545		550		555
Gly Ile Thr Lys	Gly Lys His Phe Trp	Ala Phe Arg Val Glu	Pro		
	560		565		570
Tyr Ser Tyr Leu	Val Lys Val Gly Val	Ala Ser Ser Asp Lys	Leu		
	575		580		585
Gln Glu Trp Leu	Arg Ser Pro Arg Asp	Ala Val Ser Pro Arg	Tyr		
	590		595		600
Glu Gln Asp Ser	Gly His Asp Ser Gly	Ser Glu Asp Ala Cys	Phe		
	605		610		615
Asp Ser Ser Gln	Pro Phe Thr Leu Val	Thr Ile Gly Met Gln	Lys		
	620		625		630
Phe Phe Ile Pro	Lys Ser Pro Thr Ser	Ser Asn Glu Pro Glu	Asn		
	635		640		645
Arg Val Leu Pro	Met Pro Thr Ser Ile	Gly Ile Phe Leu Asp	Cys		
	650		655		660
Asp Lys Gly Lys	Val Asn Phe Tyr Asp	Met Asp Gln Met Lys	Cys		

	665		670		675
Leu Tyr Glu Arg	Gln Val Asp Cys Ser	His Thr Leu Tyr Pro	Ala		
	680		685		690
Phe Ala Leu Met	Gly Ser Gly Gly Ile	Gln Leu Glu Glu Pro	Ile		
	695		700		705
Thr Ala Lys Tyr	Leu Glu Tyr Gln Glu	Asp Met			
	710		715		

&lt;210&gt; 50

&lt;211&gt; 503

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7507432CD1

&lt;400&gt; 50

Met Ser Glu Ala Arg	Gly Glu Pro Gly Ser	Gly Pro Glu Ala Gly
1	5	10 15
Ala Arg Phe Phe Cys	Thr Ala Gly Arg Gly	Leu Glu Pro Phe Val
	20	25 30
Met Arg Glu Val Arg	Ala Arg Leu Ala Ala	Thr Gln Val Glu Tyr
	35	40 45
Ile Ser Gly Lys Val	Phe Phe Thr Thr Cys	Ser Asp Leu Asn Met
	50	55 60
Leu Lys Lys Leu Lys	Ser Ala Glu Arg Leu	Phe Leu Leu Ile Lys
	65	70 75
Lys Gln Phe Pro Leu	Ile Ile Ser Ser Val	Ser Lys Gly Lys Ile
	80	85 90
Phe Asn Glu Met Gln	Arg Leu Ile Asn Glu	Asp Pro Gly Ser Trp
	95	100 105
Leu Asn Ala Ile Ser	Ile Trp Lys Asn Leu	Leu Glu Leu Asp Ala
	110	115 120
Lys Lys Glu Lys Leu	Ser Gln Arg Asp Asp	Asn Gln Leu Lys Arg
	125	130 135
Lys Val Gly Glu Asn	Glu Ile Ile Ala Lys	Lys Leu Lys Ile Glu
	140	145 150
Gln Met Gln Lys Ile	Glu Glu Asn Arg Asp	Cys Gln Leu Glu Lys
	155	160 165
Gln Ile Lys Glu Glu	Thr Leu Glu Gln Arg	Asp Phe Thr Thr Lys
	170	175 180
Ser Glu Lys Phe Gln	Glu Glu Glu Phe Gln	Asn Asp Ile Glu Lys
	185	190 195
Ala Ile Asp Thr His	Asn Gln Asn Asp Leu	Thr Phe Arg Val Ser
	200	205 210
Cys Arg Cys Ser Gly	Thr Ile Gly Lys Ala	Phe Thr Ala Gln Glu
	215	220 225
Val Gly Lys Val Ile	Gly Ile Ala Ile Met	Lys His Phe Gly Trp
	230	235 240
Lys Ala Asp Leu Arg	Asn Pro Gln Leu Glu	Ile Phe Ile His Leu
	245	250 255
Asn Asp Ile Tyr Ser	Val Val Gly Ile Pro	Val Phe Arg Val Ser
	260	265 270
Leu Ala Ser Arg Ala	Tyr Ile Lys Thr Ala	Gly Leu Arg Ser Thr
	275	280 285

```

Ile Ala Trp Ala Met Ala Ser Leu Ala Asp Ile Lys Ala Gly Ala
290 295 300
Phe Val Leu Asp Pro Met Cys Gly Leu Gly Thr Ile Leu Leu Glu
305 310 315
Ala Ala Lys Glu Trp Pro Asp Val Tyr Tyr Val Gly Ala Asp Val
320 325 330
Ser Asp Ser Gln Leu Leu Gly Thr Trp Asp Asn Leu Lys Ala Ala
335 340 345
Gly Leu Glu Asp Lys Ile Glu Leu Leu Lys Ile Ser Val Ile Glu
350 355 360
Leu Pro Leu Pro Ser Glu Ser Val Asp Ile Ile Ile Ser Asp Ile
365 370 375
Pro Phe Gly Lys Lys Phe Lys Leu Gly Lys Asp Ile Lys Ser Ile
380 385 390
Leu Gln Glu Met Glu Arg Val Leu His Val Gly Gly Thr Ile Val
395 400 405
Leu Leu Leu Ser Glu Asp His His Arg Arg Leu Thr Asp Cys Lys
410 415 420
Glu Ser Asn Ile Pro Phe Asn Ser Lys Asp Ser His Thr Asp Glu
425 430 435
Pro Gly Ile Lys Lys Cys Leu Asn Pro Glu Glu Lys Thr Gly Ala
440 445 450
Phe Lys Thr Ala Ser Thr Ser Phe Glu Ala Ser Asn His Lys Phe
455 460 465
Leu Asp Arg Met Ser Pro Phe Gly Ser Leu Val Pro Val Glu Cys
470 475 480
Tyr Lys Val Ser Leu Gly Lys Thr Asp Ala Phe Ile Cys Lys Tyr
485 490 495
Lys Lys Ser His Ser Ser Gly Leu
500

```

&lt;210&gt; 51

&lt;211&gt; 410

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 5639578CD1

&lt;400&gt; 51

```

Met Ser Leu Pro Glu Ser Pro His Ser Pro Ala Thr Leu Asp Tyr
1 5 10 15
Ala Leu Glu Asp Pro His Gln Gly Gln Arg Ser Arg Glu Lys Ser
20 25 30
Lys Ala Thr Glu Val Met Ala Asp Met Phe Asp Gly Arg Leu Glu
35 40 45
Pro Ile Val Phe Pro Pro Pro Arg Leu Pro Glu Glu Gly Val Ala
50 55 60
Pro Gln Asp Pro Ala Asp Gly Gly His Thr Phe His Ile Leu Val
65 70 75
Asp Ala Gly Arg Ser His Gly Ala Ile Lys Ala Gly Gln Glu Val
80 85 90
Thr Pro Pro Pro Ala Glu Gly Leu Glu Ala Ala Ser Ala Ser Leu
95 100 105
Thr Thr Asp Gly Ser Leu Lys Asn Gly Phe Pro Gly Glu Glu Thr

```

	110		115		120
His Gly Leu Gly	Gly Glu Lys Ala Leu	Glu Thr Cys Gly Ala Gly			
	125		130		135
Arg Ser Glu Ser	Glu Val Ile Ala Glu	Gly Lys Ala Glu Asp Val			
	140		145		150
Lys Pro Glu Glu	Cys Ala Met Phe Ser	Ala Pro Val Asp Glu Lys			
	155		160		165
Pro Gly Gly Glu	Glu Met Asp Val Ala	Glu Glu Asn Arg Ala Ile			
	170		175		180
Asp Glu Val Asn	Arg Glu Ala Gly Pro	Gly Pro Gly Pro Gly Pro			
	185		190		195
Leu Asn Val Gly	Leu His Leu Asn Pro	Leu Glu Ser Ile Gln Leu			
	200		205		210
Glu Leu Asp Ser	Val Asn Ala Glu Ala	Asp Arg Ala Leu Leu Gln			
	215		220		225
Val Glu Arg Arg	Phe Gly Gln Ile His	Glu Tyr Tyr Leu Glu Gln			
	230		235		240
Arg Asn Asp Ile	Ile Arg Asn Ile Pro	Gly Phe Trp Val Thr Ala			
	245		250		255
Phe Arg His His	Pro Gln Leu Ser Ala	Met Ile Arg Gly Gln Asp			
	260		265		270
Ala Glu Met Leu	Ser Tyr Leu Thr Asn	Leu Glu Val Lys Glu Leu			
	275		280		285
Arg His Pro Arg	Thr Gly Cys Lys Phe	Lys Phe Phe Phe Gln Arg			
	290		295		300
Asn Pro Tyr Phe	Arg Asn Lys Leu Ile	Val Lys Val Tyr Glu Val			
	305		310		315
Arg Ser Phe Gly	Gln Val Val Ser Phe	Ser Thr Leu Ile Met Trp			
	320		325		330
Arg Arg Gly His	Gly Pro Gln Ser Phe	Ile His Arg Asn Arg His			
	335		340		345
Val Ile Cys Ser	Phe Phe Thr Trp Phe	Ser Asp His Ser Leu Pro			
	350		355		360
Glu Ser Asp Arg	Ile Ala Gln Ile Ile	Lys Glu Asp Leu Trp Ser			
	365		370		375
Asn Pro Leu Gln	Tyr Tyr Leu Leu Gly	Glu Asp Ala His Arg Ala			
	380		385		390
Arg Arg Arg Leu	Val Arg Glu Pro Val	Glu Ile Pro Arg Pro Phe			
	395		400		405
Gly Phe Gln Cys	Gly				
	410				

&lt;210&gt; 52

&lt;211&gt; 1056

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7509080CD1

&lt;400&gt; 52

Met Glu Pro Gly	Cys Asp Glu Phe Leu	Pro Pro Pro Glu Cys Pro
1	5	10 15
Val Phe Glu Pro	Ser Trp Ala Glu Phe	Gln Asp Pro Leu Gly Tyr
	20	25 30

Ile	Ala	Lys	Ile	Arg	Pro	Ile	Ala	Glu	Lys	Ser	Gly	Ile	Cys	Lys
				35					40					45
Ile	Arg	Pro	Pro	Ala	Asp	Trp	Gln	Pro	Pro	Phe	Ala	Val	Glu	Val
				50					55					60
Asp	Asn	Phe	Arg	Phe	Thr	Pro	Arg	Val	Gln	Arg	Leu	Asn	Glu	Leu
				65					70					75
Glu	Ala	Gln	Thr	Arg	Val	Lys	Leu	Asn	Tyr	Leu	Asp	Gln	Ile	Ala
				80					85					90
Lys	Phe	Trp	Glu	Ile	Gln	Gly	Ser	Ser	Leu	Lys	Ile	Pro	Asn	Val
				95					100					105
Glu	Arg	Lys	Ile	Leu	Asp	Leu	Tyr	Ser	Leu	Ser	Lys	Gln	Cys	Asn
				110					115					120
Thr	His	Pro	Phe	Asp	Asn	Glu	Val	Lys	Asp	Lys	Glu	Tyr	Lys	Pro
				125					130					135
His	Ser	Ile	Pro	Leu	Arg	Gln	Ser	Val	Gln	Pro	Ser	Lys	Phe	Ser
				140					145					150
Ser	Tyr	Ser	Arg	Arg	Ala	Lys	Arg	Leu	Gln	Pro	Asp	Pro	Glu	Pro
				155					160					165
Thr	Glu	Glu	Asp	Ile	Glu	Lys	His	Pro	Glu	Leu	Lys	Lys	Leu	Gln
				170					175					180
Ile	Tyr	Gly	Pro	Gly	Pro	Lys	Met	Met	Gly	Leu	Gly	Leu	Met	Ala
				185					190					195
Lys	Asp	Lys	Asp	Lys	Thr	Val	His	Lys	Lys	Val	Thr	Cys	Pro	Pro
				200					205					210
Thr	Val	Thr	Val	Lys	Asp	Glu	Gln	Ser	Gly	Gly	Gly	Asn	Val	Ser
				215					220					225
Ser	Thr	Leu	Leu	Lys	Gln	His	Leu	Ser	Leu	Glu	Pro	Cys	Thr	Lys
				230					235					240
Thr	Thr	Met	Gln	Leu	Arg	Lys	Asn	His	Ser	Ser	Ala	Gln	Phe	Ile
				245					250					255
Asp	Ser	Tyr	Ile	Cys	Gln	Val	Cys	Ser	Arg	Gly	Asp	Glu	Asp	Asp
				260					265					270
Lys	Leu	Leu	Phe	Cys	Asp	Gly	Cys	Asp	Asp	Asn	Tyr	His	Ile	Phe
				275					280					285
Cys	Leu	Leu	Pro	Pro	Leu	Pro	Glu	Ile	Pro	Arg	Gly	Ile	Trp	Arg
				290					295					300
Cys	Pro	Lys	Cys	Ile	Leu	Ala	Glu	Cys	Lys	Gln	Pro	Pro	Glu	Ala
				305					310					315
Phe	Gly	Phe	Glu	Gln	Ala	Thr	Gln	Glu	Tyr	Ser	Leu	Gln	Ser	Phe
				320					325					330
Gly	Glu	Met	Ala	Asp	Ser	Phe	Lys	Ser	Asp	Tyr	Phe	Asn	Met	Pro
				335					340					345
Val	His	Met	Val	Pro	Thr	Glu	Leu	Val	Glu	Lys	Glu	Phe	Trp	Arg
				350					355					360
Leu	Val	Ser	Ser	Ile	Glu	Glu	Asp	Val	Thr	Val	Glu	Tyr	Gly	Ala
				365					370					375
Asp	Ile	His	Ser	Lys	Glu	Phe	Gly	Ser	Gly	Phe	Pro	Val	Ser	Asn
				380					385					390
Ser	Lys	Gln	Asn	Leu	Ser	Pro	Glu	Glu	Lys	Glu	Tyr	Ala	Thr	Ser
				395					400					405
Gly	Trp	Asn	Leu	Asn	Val	Met	Pro	Val	Leu	Asp	Gln	Ser	Val	Leu
				410					415					420
Cys	His	Ile	Asn	Ala	Asp	Ile	Ser	Gly	Met	Lys	Val	Pro	Trp	Leu
				425					430					435
Tyr	Val	Gly	Met	Val	Phe	Ser	Ala	Phe	Cys	Trp	His	Ile	Glu	Asp
				440					445					450

His Trp Ser Tyr	Ser Ile Asn Tyr	Leu His Trp Gly	Glu Pro Lys
455		460	465
Thr Trp Tyr Gly	Val Pro Ser Leu	Ala Ala Glu His	Leu Glu Glu
470		475	480
Val Met Lys Met	Leu Thr Pro Glu	Leu Phe Asp Ser	Gln Pro Asp
485		490	495
Leu Leu His Gln	Leu Val Thr Leu	Met Asn Pro Asn	Thr Leu Met
500		505	510
Ser His Gly Val	Pro Val Val Arg	Thr Asn Gln Cys	Ala Gly Glu
515		520	525
Phe Val Ile Thr	Phe Pro Arg Ala	Tyr His Ser Gly	Phe Asn Gln
530		535	540
Gly Tyr Asn Phe	Ala Glu Ala Val	Asn Phe Cys Thr	Ala Asp Trp
545		550	555
Leu Pro Ala Gly	Arg Gln Cys Ile	Glu His Tyr Arg	Arg Leu Arg
560		565	570
Arg Tyr Cys Val	Phe Ser His Glu	Glu Leu Ile Cys	Lys Met Ala
575		580	585
Ala Phe Pro Glu	Thr Leu Asp Leu	Asn Leu Ala Val	Ala Val His
590		595	600
Lys Glu Met Phe	Ile Met Val Gln	Glu Glu Arg Arg	Leu Arg Lys
605		610	615
Ala Leu Leu Glu	Lys Gly Val Thr	Glu Ala Glu Arg	Glu Ala Phe
620		625	630
Glu Leu Leu Pro	Asp Asp Glu Arg	Gln Cys Ile Lys	Cys Lys Thr
635		640	645
Thr Cys Phe Leu	Ser Ala Leu Ala	Cys Tyr Asp Cys	Pro Asp Gly
650		655	660
Leu Val Cys Leu	Ser His Ile Asn	Asp Leu Cys Lys	Cys Ser Ser
665		670	675
Ser Arg Gln Tyr	Leu Arg Tyr Arg	Tyr Thr Leu Asp	Glu Leu Pro
680		685	690
Thr Met Leu His	Lys Leu Lys Ile	Arg Ala Glu Ser	Phe Asp Thr
695		700	705
Trp Ala Asn Lys	Val Arg Val Ala	Leu Glu Val Glu	Asp Gly Arg
710		715	720
Lys Arg Ser Phe	Glu Glu Leu Arg	Ala Leu Glu Ser	Glu Ala Arg
725		730	735
Glu Arg Arg Phe	Pro Asn Ser Glu	Leu Leu Gln Arg	Leu Lys Asn
740		745	750
Cys Leu Ser Glu	Val Glu Ala Cys	Ile Ala Gln Val	Leu Gly Leu
755		760	765
Val Ser Gly Gln	Val Ala Arg Met	Asp Thr Pro Gln	Leu Thr Leu
770		775	780
Thr Glu Leu Arg	Val Leu Leu Glu	Gln Met Gly Ser	Leu Pro Cys
785		790	795
Ala Met His Gln	Ile Gly Asp Val	Lys Asp Val Leu	Glu Gln Val
800		805	810
Glu Ala Tyr Gln	Ala Glu Ala Arg	Glu Ala Leu Ala	Thr Leu Pro
815		820	825
Ser Ser Pro Gly	Leu Leu Arg Ser	Leu Leu Glu Arg	Gly Gln Gln
830		835	840
Leu Gly Val Glu	Val Pro Glu Ala	His Gln Leu Gln	Gln Gln Val
845		850	855
Glu Gln Ala Gln	Trp Leu Asp Glu	Val Lys Gln Ala	Leu Ala Pro
860		865	870

```

Ser Ala His Arg Gly Ser Leu Val Ile Met Gln Gly Leu Leu Val
      875                      880                      885
Met Gly Ala Lys Ile Ala Ser Ser Pro Ser Val Asp Lys Ala Arg
      890                      895                      900
Ala Glu Leu Gln Glu Leu Leu Thr Ile Ala Glu Arg Trp Glu Glu
      905                      910                      915
Lys Ala His Phe Cys Leu Glu Ala Arg Gln Lys His Pro Pro Ala
      920                      925                      930
Thr Leu Glu Ala Ile Ile Arg Glu Thr Glu Asn Ile Pro Val His
      935                      940                      945
Leu Pro Asn Ile Gln Ala Leu Lys Glu Ala Leu Thr Lys Ala Gln
      950                      955                      960
Ala Trp Ile Ala Asp Val Asp Glu Ile Gln Asn Gly Asp His Tyr
      965                      970                      975
Pro Cys Leu Asp Asp Leu Glu Gly Leu Val Ala Val Gly Arg Asp
      980                      985                      990
Leu Pro Val Gly Leu Glu Glu Leu Arg Gln Leu Glu Leu Gln Val
      995                      1000                     1005
Leu Thr Ala His Ser Trp Arg Glu Lys Ala Ser Lys Thr Phe Leu
      1010                     1015                     1020
Lys Lys Asn Ser Cys Tyr Thr Leu Leu Glu Val Arg Ser Glu Thr
      1025                     1030                     1035
Leu Thr His Ser Leu Phe Phe Ile Trp Pro Gly Cys Cys Glu Met
      1040                     1045                     1050
Ala His Ile Met Arg Thr
      1055

```

&lt;210&gt; 53

&lt;211&gt; 103

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7505899CD1

&lt;400&gt; 53

```

Met Arg Tyr Val Ala Ser Tyr Leu Leu Ala Ala Leu Gly Gly Asn
  1          5          10          15
Ser Ser Pro Ser Ala Lys Asp Ile Lys Lys Ile Leu Asp Ser Val
  20          25          30
Ile Ser Glu Leu Asn Gly Lys Asn Ile Glu Asp Val Ile Ala Gln
  35          40          45
Gly Ile Gly Lys Leu Ala Ser Val Pro Ala Gly Gly Ala Val Ala
  50          55          60
Val Ser Ala Ala Pro Gly Ser Ala Ala Pro Ala Ala Gly Ser Ala
  65          70          75
Pro Ala Ala Ala Glu Glu Lys Lys Asp Glu Lys Lys Glu Glu Ser
  80          85          90
Glu Glu Ser Asp Asp Asp Met Gly Phe Gly Leu Phe Asp
  95          100

```

&lt;210&gt; 54

&lt;211&gt; 232

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens



&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7505904CD1

&lt;400&gt; 54

```

Met Glu Tyr Pro Ala Pro Ala Thr Val Gln Ala Ala Asp Gly Gly
 1          5          10          15
Ala Ala Gly Pro Tyr Ser Ser Ser Glu Leu Leu Glu Gly Gln Glu
 20          25          30
Pro Asp Gly Val Arg Phe Asp Arg Glu Arg Ala Arg Arg Leu Trp
 35          40          45
Glu Ala Val Ser Gly Ala Gln Pro Val Gly Arg Glu Glu Val Glu
 50          55          60
His Met Ile Gln Lys Asn Gln Cys Leu Phe Thr Asn Thr Gln Cys
 65          70          75
Lys Val Cys Cys Ala Leu Leu Ile Ser Glu Ser Gln Lys Leu Ala
 80          85          90
His Tyr Gln Ser Lys Lys His Ala Asn Lys Val Lys Arg Tyr Leu
 95          100         105
Ala Ile His Gly Met Glu Thr Leu Lys Gly Glu Thr Lys Lys Leu
110          115         120
Asp Ser Asp Gln Lys Ser Ser Arg Ser Lys Asp Lys Asn Gln Cys
125          130         135
Cys Pro Ile Cys Asn Met Thr Phe Ser Ser Pro Val Val Ala Gln
140          145         150
Ser His Tyr Leu Gly Lys Thr His Ala Lys Asn Leu Lys Leu Lys
155          160         165
Gln Gln Ser Thr Lys Val Glu Ala Gly Lys Gly Tyr Pro Cys Lys
170          175         180
Thr Cys Lys Ile Val Leu Asn Ser Ile Glu Gln Tyr Gln Ala His
185          190         195
Val Ser Gly Phe Lys His Lys Asn Gln Ser Pro Lys Thr Val Ala
200          205         210
Ser Ser Leu Gly Gln Ile Pro Met Gln Arg Gln Pro Ile Gln Lys
215          220         225
Asp Ser Thr Thr Leu Glu Asp
230

```

&lt;210&gt; 55

&lt;211&gt; 508

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7509224CD1

&lt;400&gt; 55

```

Met Thr Ser Leu Met Pro Gly Ala Gly Leu Leu Pro Ile Pro Thr
 1          5          10          15
Pro Asn Pro Leu Thr Thr Leu Gly Val Ser Leu Ser Ser Leu Gly
 20          25          30
Ala Ile Pro Ala Ala Ala Leu Asp Pro Asn Ile Ala Thr Leu Gly
 35          40          45
Glu Ile Pro Gln Pro Pro Leu Met Gly Asn Val Asp Pro Ser Lys
 50          55          60

```

Ile	Asp	Glu	Ile	Arg	Arg	Thr	Val	Tyr	Val	Gly	Asn	Leu	Asn	Ser
				65					70					75
Gln	Thr	Thr	Thr	Ala	Asp	Gln	Leu	Leu	Glu	Phe	Phe	Lys	Gln	Val
				80					85					90
Gly	Glu	Val	Lys	Phe	Val	Arg	Met	Ala	Gly	Asp	Glu	Thr	Gln	Pro
				95					100					105
Thr	Arg	Phe	Ala	Phe	Val	Glu	Phe	Ala	Asp	Gln	Asn	Ser	Val	Pro
				110					115					120
Arg	Ala	Leu	Ala	Phe	Asn	Gly	Val	Met	Phe	Gly	Asp	Arg	Pro	Leu
				125					130					135
Lys	Ile	Asn	His	Ser	Asn	Asn	Ala	Ile	Val	Lys	Pro	Pro	Glu	Met
				140					145					150
Thr	Pro	Gln	Ala	Ala	Ala	Lys	Glu	Leu	Glu	Glu	Val	Met	Lys	Arg
				155					160					165
Val	Arg	Glu	Ala	Gln	Ser	Phe	Ile	Ser	Ala	Ala	Ile	Glu	Pro	Glu
				170					175					180
Ser	Gly	Lys	Ser	Asn	Glu	Arg	Lys	Gly	Gly	Arg	Ser	Arg	Ser	His
				185					190					195
Thr	Arg	Ser	Lys	Ser	Arg	Ser	Ser	Ser	Lys	Ser	His	Ser	Arg	Arg
				200					205					210
Lys	Arg	Ser	Gln	Ser	Lys	His	Arg	Ser	Arg	Ser	His	Asn	Arg	Ser
				215					220					225
Arg	Ser	Arg	Gln	Lys	Asp	Arg	Arg	Arg	Ser	Lys	Ser	Pro	His	Lys
				230					235					240
Lys	Arg	Ser	Lys	Ser	Arg	Glu	Arg	Arg	Lys	Ser	Arg	Ser	Arg	Ser
				245					250					255
His	Ser	Arg	Asp	Lys	Arg	Lys	Asp	Thr	Arg	Glu	Lys	Ile	Lys	Glu
				260					265					270
Lys	Glu	Arg	Val	Lys	Glu	Lys	Asp	Arg	Glu	Lys	Glu	Arg	Glu	Arg
				275					280					285
Glu	Lys	Glu	Arg	Glu	Lys	Glu	Lys	Glu	Arg	Gly	Lys	Asn	Lys	Asp
				290					295					300
Arg	Asp	Lys	Glu	Arg	Glu	Lys	Asp	Arg	Glu	Lys	Asp	Lys	Glu	Lys
				305					310					315
Asp	Arg	Glu	Arg	Glu	Arg	Glu	Lys	Glu	His	Glu	Lys	Asp	Arg	Asp
				320					325					330
Lys	Glu	Lys	Glu	Lys	Glu	Gln	Asp	Lys	Glu	Lys	Glu	Arg	Glu	Lys
				335					340					345
Asp	Arg	Ser	Lys	Glu	Ile	Asp	Glu	Lys	Arg	Lys	Lys	Asp	Lys	Lys
				350					355					360
Ser	Arg	Thr	Pro	Pro	Arg	Ser	Tyr	Asn	Ala	Ser	Arg	Arg	Ser	Arg
				365					370					375
Ser	Ser	Ser	Arg	Glu	Arg	Arg	Arg	Arg	Arg	Ser	Arg	Ser	Ser	Ser
				380					385					390
Arg	Ser	Pro	Arg	Thr	Ser	Lys	Thr	Ile	Lys	Arg	Lys	Ser	Ser	Arg
				395					400					405
Ser	Pro	Ser	Pro	Arg	Ser	Arg	Asn	Lys	Lys	Asp	Lys	Lys	Arg	Glu
				410					415					420
Lys	Glu	Arg	Asp	His	Ile	Ser	Glu	Arg	Arg	Glu	Arg	Glu	Arg	Ser
				425					430					435
Thr	Ser	Met	Arg	Lys	Ser	Ser	Asn	Asp	Arg	Asp	Gly	Lys	Glu	Lys
				440					445					450
Leu	Glu	Lys	Asn	Ser	Thr	Ser	Leu	Lys	Glu	Lys	Glu	His	Asn	Lys
				455					460					465
Glu	Pro	Asp	Ser	Ser	Val	Ser	Lys	Glu	Val	Asp	Asp	Lys	Asp	Ala
				470					475					480

Pro Arg Thr Glu Glu Asn Lys Ile Gln His Asn Gly Asn Cys Gln  
 485 490 495  
 Leu Asn Glu Glu Asn Leu Ser Thr Lys Thr Lys Ala Val  
 500 505

<210> 56  
 <211> 274  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7505922CD1

<400> 56  
 Met Phe Gln Ala Ala Glu Arg Pro Gln Glu Trp Ala Met Glu Gly  
 1 5 10 15  
 Pro Arg Asp Gly Leu Lys Lys Glu Arg Leu Leu Asp Asp Arg His  
 20 25 30  
 Asp Ser Gly Leu Asp Ser Met Lys Asp Glu Glu Tyr Glu Gln Met  
 35 40 45  
 Val Lys Glu Leu Gln Glu Ile Arg Leu Glu Pro Gln Glu Val Pro  
 50 55 60  
 Arg Gly Ser Glu Pro Trp Lys Gln Gln Leu Thr Glu Asp Gly Asp  
 65 70 75  
 Ser Phe Leu His Leu Ala Ile Ile His Glu Glu Lys Ala Leu Thr  
 80 85 90  
 Met Glu Val Ile Arg Gln Val Lys Gly Asp Leu Ala Phe Leu Asn  
 95 100 105  
 Phe Gln Asn Asn Leu Gln Gln Thr Pro Leu His Leu Ala Val Ile  
 110 115 120  
 Thr Asn Gln Pro Glu Ile Ala Glu Ala Leu Leu Gly Ala Gly Cys  
 125 130 135  
 Asp Pro Glu Leu Arg Asp Phe Arg Gly Asn Thr Pro Leu His Leu  
 140 145 150  
 Ala Cys Glu Gln Gly Cys Leu Ala Ser Val Gly Val Leu Thr Gln  
 155 160 165  
 Ser Cys Thr Thr Pro His Leu His Ser Ile Leu Lys Ala Thr Asn  
 170 175 180  
 Tyr Asn Gly His Thr Cys Leu His Leu Ala Ser Ile His Gly Tyr  
 185 190 195  
 Leu Gly Ile Val Glu Leu Leu Val Ser Leu Gly Ala Asp Val Asn  
 200 205 210  
 Ala Gln Leu Thr Trp Gly Arg Pro Ser Thr Arg Ile Gln Gln Gln  
 215 220 225  
 Leu Gly Gln Leu Thr Leu Glu Asn Leu Gln Met Leu Pro Glu Ser  
 230 235 240  
 Glu Asp Glu Glu Ser Tyr Asp Thr Glu Ser Glu Phe Thr Glu Phe  
 245 250 255  
 Thr Glu Asp Glu Leu Pro Tyr Asp Asp Cys Val Phe Gly Gly Gln  
 260 265 270  
 Arg Leu Thr Leu

<210> 57  
 <211> 527

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7507695CD1

&lt;400&gt; 57

```

Met Thr Ala Glu Leu Arg Glu Ala Met Ala Leu Ala Pro Trp Gly
  1          5          10          15
Pro Val Lys Val Lys Lys Glu Glu Glu Glu Glu Asn Phe Pro
          20          25          30
Gly Gln Ala Ser Ser Gln Gln Val His Ser Glu Asn Ile Lys Val
          35          40          45
Trp Ala Pro Val Gln Gly Leu Gln Thr Gly Leu Asp Gly Ser Glu
          50          55          60
Glu Glu Glu Lys Gly Gln Asn Ile Ser Trp Asp Met Ala Val Val
          65          70          75
Leu Lys Ala Thr Gln Glu Ala Pro Ala Ala Ser Thr Leu Gly Ser
          80          85          90
Tyr Ser Leu Pro Gly Thr Leu Ala Lys Ser Glu Ile Leu Glu Thr
          95          100          105
His Gly Thr Met Asn Phe Leu Gly Ala Glu Thr Lys Asn Leu Gln
          110          115          120
Leu Leu Val Pro Lys Thr Glu Ile Cys Glu Glu Ala Glu Lys Pro
          125          130          135
Leu Ile Ile Ser Glu Arg Ile Gln Lys Ala Asp Pro Gln Gly Pro
          140          145          150
Glu Leu Gly Glu Ala Cys Glu Lys Gly Asn Met Leu Lys Arg Gln
          155          160          165
Arg Ile Lys Arg Glu Lys Lys Asp Phe Arg Gln Val Ile Val Asn
          170          175          180
Asp Cys His Leu Pro Glu Ser Phe Lys Glu Glu Glu Asn Gln Lys
          185          190          195
Cys Lys Lys Ser Gly Gly Lys Tyr Ser Leu Asn Ser Gly Ala Val
          200          205          210
Lys Asn Pro Lys Thr Gln Leu Gly Gln Lys Pro Phe Thr Cys Ser
          215          220          225
Val Cys Gly Lys Gly Phe Ser Gln Ser Ala Asn Leu Val Val His
          230          235          240
Gln Arg Ile His Thr Gly Glu Lys Pro Phe Glu Cys His Glu Cys
          245          250          255
Gly Lys Ala Phe Ile Gln Ser Ala Asn Leu Val Val His Gln Arg
          260          265          270
Ile His Thr Gly Gln Lys Pro Tyr Val Cys Ser Lys Cys Gly Lys
          275          280          285
Ala Phe Thr Gln Ser Ser Asn Leu Thr Val His Gln Lys Ile His
          290          295          300
Ser Leu Glu Lys Thr Phe Lys Cys Asn Glu Cys Glu Lys Ala Phe
          305          310          315
Ser Tyr Ser Ser Gln Leu Ala Arg His Gln Lys Val His Ile Thr
          320          325          330
Glu Lys Cys Tyr Glu Cys Asn Glu Cys Gly Lys Thr Phe Thr Arg
          335          340          345
Ser Ser Asn Leu Ile Val His Gln Arg Ile His Thr Gly Glu Lys
          350          355          360

```

```

Pro Phe Ala Cys Asn Asp Cys Gly Lys Ala Phe Thr Gln Ser Ala
      365                      370                      375
Asn Leu Ile Val His Gln Arg Ser His Thr Gly Glu Lys Pro Tyr
      380                      385                      390
Glu Cys Lys Glu Cys Gly Lys Ala Phe Ser Cys Phe Ser His Leu
      395                      400                      405
Ile Val His Gln Arg Ile His Thr Ala Glu Lys Pro Tyr Asp Cys
      410                      415                      420
Ser Glu Cys Gly Lys Ala Phe Ser Gln Leu Ser Cys Leu Ile Val
      425                      430                      435
His Gln Arg Ile His Ser Gly Asp Leu Pro Tyr Val Cys Asn Glu
      440                      445                      450
Cys Gly Lys Ala Phe Thr Cys Ser Ser Tyr Leu Leu Ile His Gln
      455                      460                      465
Arg Ile His Asn Gly Glu Lys Pro Tyr Thr Cys Asn Glu Cys Gly
      470                      475                      480
Lys Ala Phe Arg Gln Arg Ser Ser Leu Thr Val His Gln Arg Thr
      485                      490                      495
His Thr Gly Glu Lys Pro Tyr Glu Cys Glu Lys Cys Gly Ala Ala
      500                      505                      510
Phe Ile Ser Asn Ser His Leu Met Arg Tyr His Arg Thr His Leu
      515                      520                      525
Val Glu

```

&lt;210&gt; 58

&lt;211&gt; 2629

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7506140CB1

&lt;400&gt; 58

```

cccttccctc ccccatacct agcccagagtc tgagccctaa cgagaaggct gggcctaggc 60
cgctggatgc tggagtgaag ggaaggaggaga aagggaaaaa gcggaagag tcgagaaggc 120
tgagtgttaa gaggccaagt gcgacgcgcg tatccgggca gacggactga cggacggggc 180
cgtgcttctg ccgcgggctgc ggcgcccgcg cgagtcgcgt ctaagcggcg gcggcgggtg 240
cagcggcgga aaccgaaggg gagccatggc ggctgacagt cgggaggaga aagatggaga 300
acttaatgtt ctggatgata ttttaactga agtaccagaa caagatgatg aactgtataa 360
tccagagagt gaacaagata aaaatgagaa aaagggatca aaaagaaaaa gtgatcgaat 420
ggaatctact gataccaaac gacaaaagcc ttctgtccat tcaagacaac tggtttctaa 480
gccactgagc tcatctgtta gcaataacaa aagaatagtt agtacaaaag gaaagtcagc 540
cacagagtat aaaaatgagg aatatcaaag atctgaaaga aacaagcgtc tagatgctga 600
tcggaaaatt cgtctatcaa gtagtgcctc cagagaacct tataagaatc aacctgaaaa 660
aacctgtgtc cggaaggagg atcctgaaag gagggccaaa tctcctacgc cagatgggtc 720
tgagagaatt gggcttgaag tggatagacg tgcaagcaga tccagccagt cttctaagga 780
agaagtgaac tctgaagaat atggctctga ccatgagact ggcagcagtg gttcttctga 840
tgagcaaggg aacaacactg agaatgagga ggaaggagtg gaagaagatg tggaggaaga 900
tgaagaggag gaggaggagg aggaagaaga agaatatgaa caggatgaga gagaccagaa 960
agaggaggga aatgattatg acactcgaag tgaggccagt gactctgggt ctgaatctgt 1020
ttccttcaca gatgggtctg tcagatctgg ttcaggcaca gatggatcag atgagaaaaa 1080
gaaggaaagg aagagagcta gaggcataac tccaattgtt tttgatagaa gtggaagctc 1140
tgcatacagag tcatatgcag gttcagaaaa gaagcatgag aaattatcat cttccgttcg 1200
tgctgtccga aaagatcaaa ccagtaaaact caaatatgtg cttcaagatg caagattttt 1260

```

```

cctcataaag agtaacaacc atgagaatgt gtctcttgcc aaagcgaagg gtgtatggtc 1320
cacgctccct gtaaatagaga agaaattaaa tcttgcatTT agatctgcaa ggagtgttat 1380
cttaatatTT tctgtcagag agagtggaaa atttcaagggt tttgcaagac tttcttcaga 1440
atcacatcac ggaggatctc ctatacactg ggtgcttcca gcaggaatga gtgctaaaat 1500
gctgggaggt gtcttttaaaa ttgactggat ttgcaggcgt gaattaccct tcactaagtc 1560
ggctcatctc accaatcctt ggaatgaaca taaaccagta aagatcggac gtgatggaca 1620
ggaaattgaa cttgaatgtg gaaccagct ttgtcttctg tttcccccg atgaaagtat 1680
tgacttgtat caggtcattc ataaaatgcg tcacaagaga agaatgcatt ctccagccccg 1740
atcacgagga cgtccatccc gtcgagaacc agtccgggat gtgggaaggc gtcgaccaga 1800
agattatgat attcataaca gcagaaagaa accaaggatt gactatcccc ctgagtttca 1860
ccagagacca ggggtatttaa aggatccacg ataccaggaa gtggacagtt tcacaaatct 1920
tattcccaac agacgatttt caggagtctg ccgagatgtg tttttaaatg ggtcctacaa 1980
tgattatgtg aggggaatttc ataacatggg accaccacca ccttggcaag gaatgcccc 2040
ttaccagga atggaacaac ctccacacca tccttactat cagcaccatg ctccacctcc 2100
tcaagctcat ccccttact caggacatca tccagtacca catgaagcaa gatacagaga 2160
taaacgagta catgattatg atatgagggt ggatgatttc cttcgtcgca cacaagctgt 2220
tgtcagtggt cggagaagta gaccccgta aagagaccgg gaacgagagc gagaccgcc 2280
tagagataac agacgagaca gagagcgaga tagaggacgt gatagagaaa gagaaagaga 2340
gcgattatgt gatcgagaca gagaccgagg ggagagaggt cgatatagaa gataatgggc 2400
ttttggaagc actgattgtt taaagataca aaaaatcttg tatttttttt ttgtgtgtgt 2460
ttacaagtag taaatttatt ttcagctgtc tgcctatgaa gttcattgtg tagaaggatt 2520
tattatgacc cctttgttc caagcatgca gtatcataag aactggaaaa actcaaattc 2580
gccaaaaatc cacagctgac agttgaattg acacttttat tggggcaga 2629

```

&lt;210&gt; 59

&lt;211&gt; 3902

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1889415CB1

&lt;400&gt; 59

```

aggcgctgcg gcagcgcgga gcttgtaaac agatccggcc gcaggtgacc atgtgaacta 60
cctgctcccc ggacgcttat tgtccttctc tcgatccgcy ccacaaagga gctcggcggt 120
cggggcgcgg agggacaggg gaacgagccg ggagagccgg ccggcgccact aagatggctg 180
aaggcgcccc gcgaggggtga gcggggggcg cggcgccagc agcggggagt cctcggggcg 240
gccggggcgg cggcccttcc ccgcgcgag aagccggagc caccatgtcg ttcgcgctgg 300
aggagacgct cgagtcggac tgggtggctg tgcggcccca tgtgttcgac gagcgcgaga 360
aacacaaatt cgtctttatt gtggcctgga acgagattga gggcaagttt gccataacct 420
gccacaaccg gacggccag aggcagagga gcggctcccg ggagcaagtg ggggcgcgag 480
ggggcgccga ggccggcgga gctgcgtccg acgggagccg cgggcccggc agccggcg 540
gcaggggtcg gcccgaggcc actgcctctg caactctggt taggagcccc gggccccggc 600
ggagctcggt ctgggaggag ggcggctctc ctcgagcac tcgcagcctt ctgggggacc 660
cgcggtcggt gagtcctggc agcaaagggg cggagagtcg tcttaggagc ccagtgcggg 720
ccaaacccat cccgggtcag aaaacatctg aagccgacga tgcggcgggg gcagccgctg 780
cagcagcccc gccggcgccc agagagggccc aggtgtcctc tgtacggata gtgagcgct 840
ctgggacgggt ctccgaggag atagaggtgc tggaaatggt gaaggaggac gaggcacctc 900
tggcgctctc ggacgcggag cagccggcgc ccgccaccga gctggagtct ccggccgaag 960
agtgcagctg ggccggactg ttttcttcc aggaacctgc cgcctgcac cagcagctgt 1020
gtcgggtgaa cctgcagttg gagcgtgccc tgcgggtgtt cccgaggaa cctcgggg 1080
tgtggactgt gctgtttggg ggcgccccg agatgaccga gcaggaaatc gacactctgt 1140
gttaccagct ccaggtctac ctgggcccag gcctggacac ctgcggtgtg aagatcctct 1200
cccaggtgct cttcaccgag accgatgatc ccgaggagta ttacgaaagc ctccagcagc 1260
tgcggcgaga gggctacgaa gaagtgtctc agcggggcag gaagcgcac caggagctct 1320

```

```

tggataagca caagaatata gagagcatgg tggagcttct ggacttgtat cagttggagg 1380
atgaagccta cagcagcctt gcagaagcta caaccgaact ctatcagtat ttactacagc 1440
cattccgaga catgagagaa cttgccatgc tacgaagaca gcagatcaag atttccatgg 1500
agaatgatta tctgggacct cgaagaattg agagtctaca aaaagaagat gctgattggc 1560
agcggaaagc tcacatggct gtactgtcta ttcaagatct tactgtcaag tactttgaaa 1620
taacagctaa agctcaaaaa gctgtgtatg atcgaatgag agctgatcag aagaaatttg 1680
gtaaagcatc atgggagcgc gctgctgaac ggatggaaaa actccagtat gcagtttcta 1740
aggaaacttt gcagatgatg agagctaaag agatatgctt ggaacagcgg aaacatgcac 1800
taaaggaaga gatgcagagt ttgagggttg gtacagaagc gatagcacga ttggatcagt 1860
tagaagctga ttattatgat ctgcaacttc agttgtatga agtacagttt gaaatcttga 1920
agtgtgaaga gttactattg acagcgcaac tagaaagcat caaaagactt atatcagaaa 1980
aaagagatga agtgggtatac tatgacactt acgaaagcat ggaggccatg ctggagaagg 2040
aagagatggc agcatctgag tacttacaga gagaagagct gcagaaactt cagcagaaaag 2100
cacgccagct ggaagcaaga cgtggacggg tttctgccaa gaaatcctac ctcagaaaata 2160
aaaaggaaat atgtattgca aaacacaatg aaatccaaca gcgcactcgg attgaagatg 2220
aatatagaac ccatcacaca gtacaactaa gagaaaaatt acatgatgaa gaagaaagaa 2280
aaagtgcctg ggtagccaa gagagacaga gaacactgga tagacttcta acatttaaac 2340
agaggtatcc tgggcaagtc atacttaaat caaccagatt acgactagct catgcaagaa 2400
gaaaagggtg agcaagtccg gttctccaag aggatcattg tgactcttta ccaagtgtgt 2460
tacaggtagc agaaaaaact gaagaggttg gagaaggaag agtcaagcgt gggccatcac 2520
agacaacaga accccagagc cttgtgcaac ttgaagatac ttcattaaca caacttgaag 2580
ccacctcatt acctctcagt ggtgttacct ctgaactgcc tcccactata tctcttccac 2640
ttttgaataa caacctcgaa ccatgttctg ttaccataaa tccactccca tcccctcttc 2700
ctccaacacc accacctccc ccacctcttc cccctccccc accaccacca cctctgcctg 2760
ttgctaagga cagtggccca gagacactgg agaaagatct gcctagaaaag gaggggaatg 2820
acaagaggat accaaagtca gccagtggcc cctcagcaca cctctttgac agcagccagc 2880
tggtcagtg acggaagaag ctcagaaaga ctgctgaagg tttgcagagg aggagagtga 2940
gtcacccat ggatgaggtg ctagcctcct tgaagcgtgg tagttttcat ctgaaaaagg 3000
ttgaacagcg aactctgcct ccttttccct atgaagatga tagtaataat atcttggcac 3060
aaataaggaa aggggtaaaa ttgaagaagg tacagaagga tgttttgaga gaatccttca 3120
cacttctacc cgatacagac cctctaacac ggagcatcca tgaagctctt agaagaatta 3180
aagaagcatc cccagagtca gaggacgaag aggaggcttt acctgcaca gactgggaga 3240
actaacaagt gacataacag aagaaaaata tcattcaaga ttgggtctga ttcttttggg 3300
aaatgacagt ttaagcactt ggttcctcag ttggataaca ttagatccaa agtgcatctg 3360
ttcagtgggt tgaagaaagg aagcacaatt ggcagggttat cactttccag tcgttccaat 3420
agatgatggt taacatgaat ttacatgtg caatgttcct gactgcaatg aagaaaaggc 3480
cttctggaga tttctgtttt ttaagcacat tctccttcca tccacatatt tacatccttg 3540
taatggccag ttaagctagg gcttgcctga agttggaaga acactgggtt gacagagatc 3600
tactgtgagc tgtattggga ctgctttgag aaccatcttt caatgcactg aaaagtcac 3660
tgaaaaaata gttcttcca taccagctta tattaatac tgtgacactc tgagaaagtt 3720
tatcatcatg atagctgtaa gtcaggcatt aaaatcaaat ggaaatacac ctaaaatgtc 3780
cgcttactc atttccaaag gatttttcat aaagatgttc agcaacatct gaaaagctgc 3840
atttatggag atacatctat taatatgttc acaacatgat acatttaaca tggcatattt 3900
tt

```

&lt;210&gt; 60

&lt;211&gt; 2993

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7506047CB1

&lt;400&gt; 60

cgctcgctcg ctgtcgacac gactcaccgt cccttgacca attatcatat tcacgaccc 60

```

gcaagatattc accgtgtgtg cactcgcgtg ttttcctctc tctgccgggg gaaaaaaaaag 120
agagagagag agatagagag agagagagag agagagagag agaggctcgg tcccaactgct 180
ccctacaccg cgggtccggg attcttgagc tgtgccagc tgacgagctt ttgaagatgg 240
cacaataacc gtccagtgtat gcctgacat gacagcacag ccctcttaag ccggcaaacc 300
aagaggagaa gagttgacat tggagtgaag aggacggtag ggacagcatc tgcatttttt 360
gctaaggcaa gagcaacgtt ttttagtgcc atgaatcccc aaggttctga gcaggatgtt 420
gagtattcag tgggtcagca tgcagatggg gaaaagtcaa atgtactccg caagctgttg 480
aaaaataaca tgaacaaaaa tgggtggcag gagcccagtt tccaagccag cgggtctctt 540
agtacaggct ccgaagtaca tcaggaggat atatgcagca actcttcaag agacagcccc 600
ccagagtgtc tttccctttt tggcaggcct actatgagcc agtttgatat ggatcgctta 660
tgtgatgagc acctgagagc aaagcgcgcc cgggttgaga atataattcg gggatgagc 720
cattccccca gtgtggcatt aaggggcaat gaaaatgaaa gagagatggc cccgcagctc 780
gtgagtcccc gagaaagtta cagagaaaac aaacgcaagc aaaagcttcc ccagcagcag 840
caacagagtt tccagcagct ggtttcagcc cgaaaagaac agaagcgaga ggagcgccga 900
cagctgaaac agcagctgga ggacatgcag aaacagctgc gccagctgca ggaaaagttc 960
taccaaatct atgacagcac tgattcggaa aatgatgaag atggtaacct gtctgaagac 1020
agcatgcgct cggagatcct ggatgccagg gccaggact ctgtcggag gtcagataat 1080
gagatgtgcg agctagacct aggacagttt attgaccgag ctcgagccct gatcagagag 1140
caggaaaatgg ctgaaaacaa gccgaagcga gaaggcaaca acaaaagaaag agaccatggg 1200
ccaaactcct tacaaccgga aggcaaacat ttggctgaga ccttgaacaa ggaactgaac 1260
actgccatgt cgcaagttgt ggacactgtg gtcaaagtct tttcgccaa gccctccgcg 1320
caggttctctc aggtcttccc acctctccag atccccagg ccagatttgc agtcaatggg 1380
gaaaaccaca atttccacac cgcaaccag cgcctgcagt gctttggcga cgtcatcatt 1440
ccgaaccccc tggacacctt tggcaatgtg cagatggcca gttccactga ccagacagaa 1500
gcaactgccc tgggtgtccg caaaaactcc tctgaccagt ctgcctccgg ccctgccgct 1560
ggcggccacc accagcccct gcaccagtcg cctctctctg ccaccacggg cttcaccacg 1620
tccaccttcc gccaccctt ccccttccc ttgatggcct atccatttca gagccatta 1680
gggtctccct cgggtcctt ctctggaaaa gacagagcct ctctgaatc cttagactta 1740
actagggata ccacagctc gaggaccaag atgtcatctc accactgag ccaccacct 1800
tgatccag caccaccgcc cagcaccgcc gaagggtctc ccttgtcgct cataaagtcc 1860
gagtgcggcg atcttcaaga tatgtctgaa atatcacctt attcgggaag tgcaatgcag 1920
gaaggattgt caccatca cttgaaaaaa gcaaagctca tgttttttta taccggttat 1980
cccagctcca atatgtgaa gacctactt tccgacgtaa agttcaacag atgcattacc 2040
tctcagctca tcaagtgtt tagcaattt cgtgagttt actacattca gatggagaag 2100
tacgcacgtc aagccatcaa cgatggggtc accagtactg aagagctgtc tataaccaga 2160
gactgtgagc tgtacagggc tctgaacatg cactacaata aagcaaatga ctttgagggt 2220
ccagagagat tccgtggaag tgctcagatc acattacggg agtttttcaa tgccattatc 2280
gcaggcaaag atgttgatcc ttcctggaag aaggccatat acaaggctat ctgcaagctg 2340
gatagtgaag tccctgagat tttcaaatcc ccgaactgcc tacaagagct gcttcagtag 2400
tagaaatttc aacaactctt tttgaatgta tgaagagtag cagtcccctt tggatgtcca 2460
agttatatgt gtctagattt tgatttcata tatatgtgta tgggaggcat ggatatgtta 2520
tgaaatcagc tggtaattcc tctcatcac gtttctctca ttttctttt ttttccattg 2580
caaggggatg gttgttttct tctgccttt agtttgcttt tgcccaaggc ccttaacatt 2640
tggacactta aaatagggtt aattttcagg gaaaaagaat gttggcgtgt gtaaagtctc 2700
tattagcaat gaagggaatt tgtaacgat gcattcactt gattgatgac ttattgcaaa 2760
tggcgggttg ctgaggaaaa ccatgacac agcacaactc tacagacagt gatgtgtctc 2820
ttgtttctac tgctaagaag gtctgaaaat ttaatgaaac cacttcatac atttaagtat 2880
ttgttttgtt ttgaaactca tcagtagctt ttccttacat gtttaaaaaa aattccaatg 2940
acagatgagc agctcactta tccaaagacc ccatttttctg agcgacatcg gtg 2993

```

&lt;210&gt; 61

&lt;211&gt; 679

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;



&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7505849CB1

&lt;400&gt; 61

```
ccttgacctc tggcccaaga tggtagcgcc cagagcttcg ctcttgctgc tccccgagg 60
tgaactgaag ccagcagccc cgcacatgt caaagctcgg ccgggcccgc cggggcctca 120
ggaagcccgga gagaggcgtt tccatcaacc agttttgcaa ggagttcaat gagaggacaa 180
aggacatcaa ggaaggcatt cctctgccta ccaagatttt agtgaagcct gacaggacat 240
ttgaaattaa gattggacag cccactgttt cctacttcct gaaggcagca gctgggattg 300
aaaagggggc ccggcaaaaca gggaaagagg tggcaggcct ggtgacctg aagcatgtgt 360
atgagattgc ccgcatcaaa gctcaggatg aggcatttgc cctgcaggat gtacccctgt 420
cgtctgttgt ccgctccatc atcgggtctg cccgttctct gggcattcgc gtggtgaagg 480
acctcagttc agaagagctt gcagctttcc agaaggaacg agccatcttc ctggctgtct 540
agaaggaggc agatttggct gccaagaag aagctgcca gaagtgacct ttccccacc 600
aactcccaga tttcaaagga ggtagttgca aaagctgtgc ccaaggggag gaaggaggtc 660
acaccaatat gatgatggt                                     679
```

&lt;210&gt; 62

&lt;211&gt; 918

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7505972CB1

&lt;400&gt; 62

```
gcgttaggag gctcgggtcg ttgtggtgcg ctgtcttccc gcttgcgtea gggacctgcc 60
cgactcagtg gccgccatgg catcagatga aggcaaaact tttgttgagg ggctgagttt 120
tgacaccaat gagcagtcgc tggagcaggc cttctcaaag tacggacaga tctctgaagt 180
ggtggttgtg aaagacaggg agaccagag atctcgggga tttgggtttg tcacctttga 240
gaacattgac gacgctaagg atgccatgat ggccatgaat ggggaagtctg tagatggacg 300
gcagatccga ggacggggcc gtgggttctc tagaggagga ggggaccgag gctatggggg 360
gaaccgggtc gagtcagga gtgggggcta cggaggctcc agagactact atagcagccg 420
gagtcagagt ggtgggtaca gtgaccggag ctccggcggg tcctacagag acagttacga 480
cagttacgct acacacaacg agtaaaaacc cttcctgtct aagatcgccc ttccaatggc 540
tgtgtgttta aagattgtgg gagcttcgct gaacgttaat gtgtagtaaa tgcacctcct 600
tgtattccca ctttcgtagt catttcgggt ctgatcttgt caaaccagc ctgaccgctt 660
ctgacgccgg gatggcctcg ttactagact ttctttttta aggaagtgtc gttttttttt 720
gaggggtttt aaaacatttt gaaaagcatt tacttttttt agccaggagc catgagtttt 780
tccaaaaaat cgggggggtt tgtgggggtt tttgggggtt tggtttttagt tttttgggtt 840
gcgttggcct tttttttttt tagtgggggt tgggtcccat gaagtggggg gccccactc 900
acttctctga aggacgga                                     918
```

&lt;210&gt; 63

&lt;211&gt; 2010

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7505991CB1

&lt;400&gt; 63

```
cggagacgcg aggaggggag cagagaacgc gctcgcgtag aggaattccg ggtcctcctc 60
ctcgccttac cttggtttac ctgcagccgc ctagtccctc tctccttctc tttgatgatt 120
```

```

cagggcctcg gcttccctca cgatattgca gaaagacaca gctttcctct tcctctccaa 180
accacctcga agcagggtta aggcataaatt ggaatatcat ttggagaaag tgtcatggaa 240
gttctgcgtc cacagcttat aagaattgat ggccggaatt acaggaagaa tccagtccaa 300
gaacagacct atcaacatga agaagatgaa gaggacttct atcaaggctc catggagtgt 360
gctgatgagc cctgtgatgc ctacgagggtg gagcagagccc cacaaggatt ccggtctact 420
ttgagggccc ccagcttgct ctataagcat atagtggaa agagagggga cactaggaag 480
aaaatagaaa tggagaccaa aacttctatt agcattccta aacctggaca agacgggaa 540
attgtaataca ctggccagca tcgaaatggt gtaatttcag cccgaacacg gattgatgtt 600
cttttggaac cttttcgaag aaagcagccc ttactcact tccttgccct tttcctcaat 660
gaagttgagg ttccaggaag attcctgaga ttccaggagg aagtactggc gaagtgtctc 720
atggatcatg ggggtgacag cagcattttc cagaatccta aaaagcttca tctaactatt 780
gggatgttgg tgcttttgag tgaggaagag atccagcaga catgtgagat gctacagcag 840
tgtaaagagg aattcattaa gctacaagaa ttagttgatc gagtgtctga acgttttcag 900
gcatctggac taatagtgaag agagtggaaat agtgtgaaac tgcagtctac agttatgaat 960
acactattca ggaaagaccc caatgctgaa ggcaggatca atctctacac agcggaaagg 1020
aaatatatct tcaaggaaag agaatacatt gatggccgaa atatttttaa gttgtttgag 1080
aacttctact ttggctccct aaagctgaat tcaattcaca tctctcagag gttcaccgta 1140
gacagctttg gaaactacgc ttctgtgga caaattgact tctctgagg tggatcttgg 1200
aaagcactag aaactaaaca tcttcaccag gtgctgaaga aaagtgtctt cgttttaatt 1260
gccaagcagg gatgtggaca tttggatggt gactttctct ggtggttccc catagattca 1320
ccattgcctc taatggtgtc tacacccgtc atactaccag ctgagatggt ggtgggcata 1380
aggagaattt gtgcctataa cccttagtgt gttctggttt tttttctttt aatttttaaa 1440
ttgtcgtaaa atactcataa aacatactgt cttcaccatt ttttaagtga cagttcagta 1500
acgttaactg ttaatacatt cataatgctg tgtggccgtc accgccgtcc atctccatag 1560
gcttctcagc ttgtaaaaat gaaactgtac ccattaaaca gtaattccca ctctcccgag 1620
ccccgcgagc caccattctg ctttctgtct ctctggtttt gactattctc agtatctcat 1680
ataagtggaa tcatacagtg acttgtcttt ttgtgactgg cttatttcac ttagcataat 1740
gtcctcaagg ttcattccatg ttgtgtcagg tgacaggatt tccttctgtt attatacata 1800
taatggaatg ttccgttaca tgtgcgtacc acactctgtt cctccatcag tgaacacttg 1860
gcttgcttcc tcttgactat tgggagtagt cctaaatacag acacgggggt gcaaatatct 1920
ctttgagact ctgcctttaa ttcctttacc acccagagtt ttatctttag caaaataacc 1980
attaaagttg tttgcctttt aaaaaaaaaa 2010

```

&lt;210&gt; 64

&lt;211&gt; 4677

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7506003CB1

&lt;400&gt; 64

```

gagaggatgc cgcagatgtg ttagcggcga gtccagaagc agccccagga ggtgctgggg 60
gcatcgtttc tctaactctg cctcccaggt gccaaaggagg cgtcccgga gcggtcatca 120
tggtgaagga gcagttccgg gagacggatg tggccaagaa aataagccac atctgttttg 180
gaatgaagtc acctgaggag atgcgccagc aggcgcacat ccaagttgtg agtaagaacc 240
tgtacagcca ggacaacca catgccccct tgctatatgg ggtgctcgac cataggatgg 300
gtacgagtga gaaggatcgt ccatgtgaaa cctgtgggaa aaacttggct gactgtctag 360
gccactatgg gtatatcgac ctggagttgc cgtgttttca tgtagggtac ttcagagcag 420
tcataggcat cttacagatg atctgcaaaa cctgctgcca catcatgctg tcccaagagg 480
agaagaagca gtttctggac tatctaaaga ggcccgccct gacctacctt cagaagcgag 540
gactgaaaaa gaaaatctct gacaagtgcc ggaagaaaaa catctgccat cactgtggcg 600
cttttaaatg taccgtaaag aagtgtggac tgctgaaaat aattcatgag aaatacaaga 660
ccaacaaaaa agtgggtggat ccatttgtat caaatttcct tcagtctttt gaaacagcca 720
ttgaacataa taaagaagtg gaggctctgc tgggaagggc acaggaaaaa ttgaatccct 780

```

```

tagtagttct gaatttattt aaacgaatcc cagctgaaga tgttcctcta cttctgatga 840
accagaagc cggaagccg tctgatttga ttctcacacg acttttgggtg cctcctttgt 900
gtttcagacc ctccgtttgt agtgatttga agtctggcac caatgaagat gatctgacaa 960
tgaaactgac agaaatcatt ttctaaacg atgttattaa aaagcatcgg atctcaggag 1020
ccaagaccca gatgatcatg gaggactggg atttcttgca gctgcagtgt gccctctaca 1080
ttaacagtga gctctcgggc attccctca acatggcacc caagaagtgg accagaggct 1140
tcgtccaacg cctgaaggga aaacaggggc gatttagagg aaatctctca ggaaagagag 1200
tggttttttc tggcagaaca gtcatctcgc ccgaccccaa cctccggatt gatgaggtag 1260
ctgtgccagt tcatgtggcc aaaattctaa cttttctga gaaggtaaac aaagcaaaca 1320
tcaattttct gaggaaactg gttcaaacg gccctgaggt tcaccagga gcaaacttca 1380
ttcagcagag acatacgag atgaaaaggt ttttgaaata cggaatcga gaaaagatgg 1440
ctcaagagct caagtatggt gacatcgtag agagacacct catcgatgga gatgtggtgc 1500
tgtcaatcg gcagccctcg ctgcacaaat tgagcattat ggctcatctg gccagggtca 1560
agccccaccg gaccttcaga tttaatgagt gtgtctgtac accctataat gctgactttg 1620
atgggtgatga aatgaacctt catcttcttc aaacagaaga agctaaagca gaggcccttg 1680
ttctgatggg gactaaagca aatcttgtaa ccccgaggaa tggggaaccg ctgattgtcg 1740
ctattcagga ttttctaaca ggtgcctatc tcctcactct caaggacact ttctttgatc 1800
gagccaaggg ttgccaatc attgttcaa tactgggtgg caaggatgag aaaattaaag 1860
ttcgctctcc accgctaca atcctaaagc ctgtcaccct gtggacggga aagcagatct 1920
tcagtgatcat cctcagccct agcagtgaca atccagttag ggccaacctg cgaaccaagg 1980
gcaagcagta ctgtggcaaa ggggaagatc tctgtgcaa tgattcctat gttacaatcc 2040
agaacagtga gttgatgagt ggcagcatgg acaaaggaa cctaggggtca ggatccaaga 2100
acaatatttt ttacattttg ctgcgagact ggggacagaa tgaagctgca gatgccatgt 2160
cacggctcgc caggctggct cctgtctacc tgtctaaccg tggtttctca attgggatcg 2220
gtgatgtcac acctggccaa ggactgctga aggccaagta tgagttgctg aatgccggct 2280
acaagaaatg tgatgagtac atcgaagccc tgaacacggg caagctgcag cagcagcctg 2340
gctgcactgc tgaggagacc ctggaggcac tgatcctgaa ggagctgtct gtgatccgtg 2400
accacgctgg cagtgcctgc ctccgggagc tggacaagag caacagcccc ctcaccattg 2460
ctctgtgcgg ctccaaaggt tccttcatta acatatcaca gatgattgac tgtgtgggac 2520
agcaggccat cctgggctct cgagtgccag acggctttga aaacagggtcc ttgcctcatt 2580
ttgaaaaaca ctcaaagctc ccagctgcca aaggctttgt ggctaatagc ttttattccg 2640
gtttgacacc aactgagttt ttcttcaca caatggccgg ccgggaagggt ctagtgcaca 2700
cggctgtaaa gacagctgaa acgggataca tgcagcgaag gcttgtaaaa tctcttgaag 2760
atctttgtct ccagtatgat ctgacagtec gaagctctac tggcgatatt atccagttca 2820
tttatggagg agatggctta gatcctgcag ctatggaggg aaaagatgaa cctttggagt 2880
ttaaagggtt tctggacaac atcaaagcag tcttcccggtg tcccagtgag cctgtctctca 2940
gcaaaaacga gctgatcctg accacagagt ccatcatgaa gaagagttag ttctctgtct 3000
gccaggacag ctctctgcag gaaataaaaa aattcattaa ggggtctctt gagaagatca 3060
agaaaaccag agataaatat ggcatcaatg ataacggcac aacagagccc cgtgtgctgt 3120
accagctgga ccgcatcacc cccaccaag tagaaaagtt tctggagacc tgtagggaca 3180
agtacatgag ggcacagatg gagccaggtt ctgcagtggt tgctctgtgt gccagagca 3240
ttggtgagcc aggcaccag atgacctga agactttcca ctttgcaggt gtggcctcca 3300
tgaacatcac cctgggctgt ccccgatta aagagatcat caacgcttc aaggccatca 3360
gcactccaat tatcacagca cagctagaca aggatgacga cgcgattat gctcgcctcg 3420
tgaaaggagg aattgagaaa accctcttgg gagaggtgaa cgctgagaca gtgagatatt 3480
ccatctgcac atccaagctc cgtgtgaagc ccggtgatgt ggctgttcat ggtgaggctg 3540
tggtgtgtgt caccgccaga gagaacagca agagctccat gtactacgtg ctgcagttcc 3600
tgaaagagga tctccccaag gtggtggtgc agggcattcc agaggtgtcc agagctgtca 3660
tccacattga cgagcagagt ggaaaggaga agtacaagct tctggtggaa ggtgataacc 3720
tgcgggcagc catggccaca cagggtgtga agggcaccgc aaccacctcc aataaacctc 3780
atgaggtgga gaaaactctg ggcacagagg ccgcccggac aacgatcatc aatgaaatcc 3840
agtacaccat ggtgaaccac ggcacagagg tcgacaggag gcacgtgatg ctgctctccg 3900
acctcatgac ctacaagggt gaagtcctgg gcatcactag gtttggcctg gccaaagtga 3960
aggagagtgt gctgatgctg gcctcctttg agaagacggc tgaccatctc tttgacgctg 4020
cctacttcgg gcagaaggac tctgtgtgtg ggggtgtctga gtgcatcatc atgggaatcc 4080
caatgaacat tggaaccggg ctcttcaagc tgcttcacaa ggctgacagg gacccgaacc 4140

```

```

ctcccaagag gccctgac ttcgacacaa atgaattcca catccccctt gtcacatagt 4200
ccaaagaaag aggggaccat gcctgacctt gactccttgt cctgtctcca gctgatgtat 4260
aaagagtttt gtgctccctg ggacgggggt cctgaggtcc ccacctatgc cagcaatcag 4320
agaagccctc ttggcatccc caggagcagc ttctcctctg ataggggtgca gctcacacca 4380
gtgaccctga ctgtgccacg ctgctcggga gagctgaggg ttttattgtt tgcttgcttg 4440
aaacctaatc tatagacggc cccacagctc gtgcacacac tgcttccctg gacttaaagc 4500
ccagccaggt gtcattggtc atcccagctc actgcataca tccatcggct ccccatgggt 4560
tcttcacacc tgacgatgag ccaggcctga gccccacaca ggccagggca cattcttggg 4620
ttttccattc cttgggtcatg ctggaatctc tcaatgtgac atacttatgt aaatatc 4677

```

&lt;210&gt; 65

&lt;211&gt; 1110

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 6483977CB1

&lt;400&gt; 65

```

cggctcgagc cgggactctt ggcggttgaa ggtgtgtgtc agcttttgcg tcaactcgagc 60
cctgggcgct gcttgctaaa gagccgagca cgcggtctg tcatcatgtc gcgttacggg 120
cggtagcgag gagaaaccaa ggtgtatgtt ggtaacctgg gaactggcgc tggcaaagga 180
gagttagaaa gggctttcag ttattatggt cctttaagaa ctgtatggat tgcgagaaat 240
cctccaggat ttgcctttgt ggaattcgaa gatcctagag atgcagaaga tgcagtacga 300
ggaactggatg gaaagggtgat ttgtgggtcc cgagttaggg ttgaactatc gacaggcatg 360
cctcggagat cacgttttga tagaccacct gcccgacgtc cctttgatcc aaatgataga 420
tgctatgagt gtggcgaaaa gggacattat gcttatgatt gtcacgtta cagccggcga 480
agaagaagca ggtcacggtc tagatcacat tctcgatcca gaggaaggcg atactctcgc 540
tcacgcagca ggagcagggg acgaagggtc aggtcagcat ctctcgcagc atcaagatct 600
atctctcttc gtagatcaag atcagcttca ctcaagaagat ctaggctctg ttctataaaa 660
ggatcgaggt atttccaatc cccgtcgagg tcaagatcaa gatccagggtc tatttcacga 720
ccaagaagca gtcgttcccc atcaggaagt cctcgcagaa gtgcaagtcc tgaaagaatg 780
gactgaagct ctcaagttca cccttaggg aaaagttatt ttgtttacat tattataagg 840
gatttgtgat gtctgtaaag tgtaacctag gaaagataat tcaaccatct aatcaaaatg 900
gatctggatt actatgtaaa ttcacagcag taagataata taaattttgt tgaatgtatt 960
aacatcatat ggtctgaaaa tgtgggtttt tatttggcac atttaataa aatgtttcta 1020
actagatttt tgatttgtgt tcaatattaa cacttcttaa tttgatatat ttgagagtca 1080
gacattataa ttgttaacct tattcatata 1110

```

&lt;210&gt; 66

&lt;211&gt; 687

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 6301777CB1

&lt;220&gt;

&lt;221&gt; unsure

&lt;222&gt; (1) ... (687)

&lt;223&gt; a, t, c, g, or other

&lt;400&gt; 66

```

tttttgtgtt gtgttttctc ttttcttttg ttgttctctg gtttgcttgc cttcagtcctc 60

```

```

ttttccgaaa aggggatgtgc tgcaaggcga tttacttggg aacgccaggg ttttccag 120
cacgacgttg taaaacgacg gccagtgaat tgaatttagg tgacactatt agaagannnn 180
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnngccc gggccgcgcc 240
cgggacacctca ttccggacct cggtttccct ttgctggggg ctgcagctg ccaccgcttc 300
cccgcggagc atggggctgc ctgcggggc aggggacgcg gcggagctgc gcaagagcct 360
gaagccgctg ctggagaagc gccggcgcg gcgcatcaac cagagcctga gccagcttaa 420
ggggctcatc ctgcccgtgc tgggcccggg ggatgcttct ggctggcaca cctggcttcc 480
cctccatgct cagaactgcc tcctactcta catccaggct cctgagcagc cccagctta 540
agcttcagca gctccaggg aaatccatct cggcccagag ctgactccgg atctgcccc 600
gtgacaataa caagataaat gccttccatt ctcttcgtac ttacaggatg taatccacat 660
atttacacag tgtttcattt gcggtaa 687

```

&lt;210&gt; 67

&lt;211&gt; 599

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7505976CB1

&lt;400&gt; 67

```

gacttttcgt tccggctgcc gcaggcgctt cgctgggtgca gacgcagtgc tgagcacaca 60
gctaccggac aaagagtgc gcccgagct ggagttatgg cggctacgga gccgatcttg 120
gcggccactg ggagtcgcc gcgggtgcca cggagaaac tggaggagc cggttcgagc 180
tcagcccctg agcgtaactg tgtgggctcc tcgctgccag aggcctcacc gcctgcccct 240
gagccttcca gtcccaacgc cgcggtccct gaagccatcc ctacgccccg agctgcggcc 300
tccgcggccc tggagctgcc tctcggggcc gcaccctga gcgtagcgcc tcaggccgaa 360
gctgaagcgc gctccacacc agggcccgc ggctctagac tcgggtccga gacgttccgc 420
cagcgtttcc ggcagttccg ccgcccgcag gatgtgcga tcaactggctg agcgggtggag 480
ctgcggggcg ccaggggcgg gcgctctgtg cggactgggg ccatgatcgg gccggggggc 540
ctgagcctgg gacccacccc cgtgttaatg aaaaatgagt tttggcagcg aaaaaaaaa 599

```

&lt;210&gt; 68

&lt;211&gt; 2095

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7506016CB1

&lt;400&gt; 68

```

cgccgcccgg ccggtgctcc ctggcgagc cacgcaggcg caccgcagac agacccctct 60
gccatgaacc agtccatccc agtggctccc accccacccc gccgcgtgcg gctgaagccc 120
tggctggtgg ccaggtgaa cagctgccag taccaggggc ttcaatgggt caacggggaa 180
aagaaattat tctgcatccc ctggaggcat gccacaaggc atggtcccag ccaggacgga 240
gataacacca tcttcaaggc ctgggccaag gagacaggga aatacaccga aggcgtggat 300
gaagccgatc cggccaagtga gaaggccaac ctgcgctgtg cccttaacaa gagccgggac 360
ttccgcctca tctacgacgg gcccggggac atgccacctc agccctacaa gatctacgag 420
gtctgctcca atggccctgc tcccacagac tcccagcccc ctgaggatta ctcttttggg 480
gcaggagagg aggaggaaga agaggaagag ctgcagagga tgttgccaag cctgagcctc 540
acagaggatg tcaagtggcc gccactctg cagccgcccc ctctgcagcc gcccggtggg 600
ctgggtcccc ctgctccaga cccagcccc ctggctcttc cccctggcaa ccctgctggc 660
ttcaggggagc ttctctctga ggtcctggag cctggggccc tgctgcccag cctgccccct 720
gcaggcgaac agctcctgcc agacctgctg atcagccccc acatgctgcc tctgaccgac 780

```

```

ctggagatca agtttcagta ccggggggcg ccaccccggg ccctcaccat cagcaacccc 840
catggctgcc ggctcttcta cagccagctg gaggccaccc aggagcaggt ggaactcttc 900
ggccccataa gcctggagca agtgcgcttc ccagccctg aggacatccc cagtgaacag 960
cagcgcttct acacgaacca gctgctggat gtcctggacc gcgggctcat cctccagcta 1020
cagggccagg acctttatgc catccgcctg tgtcagtga aggtgttctg gagcgggcct 1080
tgtgcctcag cccatgactc atgccccaac cccatccagc gggagggtcaa gaccaagctt 1140
ttcagcctgg agcattttct caatgagctc atcctgttcc aaaagggcca gaccaacacc 1200
ccaccacct tcgagatctt cttctgcttt ggggaagaat ggctgaccg caaacccga 1260
gagaagaagc tcattactgt acagggtgtg cctgtagcag ctgactgct gctggagatg 1320
ttctcagggg agctatcttg gtcagctgat agtatccggc tacagatctc aaaccagac 1380
ctcaaagacc gcatggtgga gcaattcaag gagctccatc acatctggca gtcccagcag 1440
cggttgcagc ctgtggccca ggccctcct ggagcaggcc ttggtgttg ccagggggcc 1500
tggcctatgc acccagctgg catgcaataa caaggctgca gacggtgact ggccctggct 1560
tcctgggtgg cggtgccggac tgatgtggag atgtgacagc cccgatgagc acctggctgg 1620
ctgcagggtc ctacctctgg gtttcctgga agtggatttg ggccaagaag gagagggaga 1680
aaggccccag cccctgcctt cccgggcctt tctctcctgg gctgtctctg gtctggtcag 1740
cctggctctc gggaattca gccatgagca gggaaagaac tctcccaacc ctggggccta 1800
gctgtatagg aggaattgac taagggtggc cactcttgt gattgcccc tttcctctg 1860
caacaaaagc cagagtgttg tggcccaagt ccccccacag ggctctgca gggcatggcc 1920
ctgatttccc tggtttgaga ctacttcct catctccctg tcctctgaga taatatgagt 1980
gagcacttag gtatcatatc agatgctcaa ggctggcagc taccctctt ttgagagtcc 2040
aagaacctgg agcagaaata atttttatgt attttggtt aataaatgtt aaaaa 2095

```

&lt;210&gt; 69

&lt;211&gt; 2843

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7506086CB1

&lt;400&gt; 69

```

actagttctc tctctctctc taggcagcaa ggagaagcag gctgacttcc tggacaagca 60
gatgagccgg gatgagcacc gagcccgagc catgaagatc cctttcacca atgacaaaat 120
catcaacctg cctgtggagg agttcaatga actgctgtcc aaataccagt tgagtgaagc 180
ccagctgagc ctcatccgag agtgggcatg gcgacttggc ctgagccgga ctcggttttc 240
tctggaaacc cccctggtaa gtgtggagga ggccgggacac tctgacccaa gacgaaaggc 300
ctgtagctcc agccaaagaa aataaacctt aggagggaga aggaaaaaaa aaatccatca 360
gctgttctcg agaacagcct gcattggaat ctacagagag gacaactaat gtgagtgaag 420
aagtgactgt atgtggactg tggagaaagt aagtcacgtg ggcccttgag gacctggact 480
gggttaggaa cagttgtact ttcagagggt aggtgtcgag aagggaaggt gaatgtggtc 540
tggagtgtgt ccttggcctt ggctccacag ggtgtgcttt cctctggggc cgtcaggggag 600
ctcatccctt gtgttctgcc aggttggggg acgggtttg aactgagga gggtaacctg 660
ctggctggag cggcagagca gtggccttga tttgtctttt ggaagatttt aaaaaccaa 720
aagcataaac attctggtcc ttcagcaatg ctttctctga agaaatactt aacggaagga 780
cttctccagt tcaccattct gctgagtttg attgggttac ggggtggacgt ggatacttac 840
ctgacctcac agcttcccc actccgggag atcatcctgg ggcccagttc tgcctatact 900
cagacctcag tccacaacct gaggaatacc ttggatggct atggtatcca cccaagagc 960
atagacctgg acaattactt cactgcccgg cggtctctca gtcaggtgag ggccctggac 1020
aggttccagg tgccaaccac tgaggtaa atgcctggctg ttcaccgaga cccagagggg 1080
tctgtctctg gcaagcagcc caactcagcc ctgcctctcg agagtccag tggcctccaa 1140
gatgtgacag gccagacaa cgggtgagca gaaagcgaaa cggagcaggg attcgggtga 1200
gatttggagg atttgggggc tgtagcccc ccagtcagtg gagacttaac caaagaggac 1260
atagatctga ttgacatcct ttggcgacag gatattgatc tgggggctgg gcgtgagggt 1320
tttgactata gtcaccgcca gaaggagcag gatgtggaga aggagctgcg agatggaggc 1380

```

```

gagcaggaca cctgggcagg cgagggcgcg gaagctctgg cacggaacct gctagtggat 1440
ggagagactg gggagagctt ccctgcacag gtgcctagt gggaggacca gacggccctg 1500
tccctggaag agtgcccttag gctgctggaa gccacctgcc cctttgggga gaatgctgag 1560
tttccagcag acatttccag cataacagaa gcagtgccta gtgagagtga gccccctgct 1620
cttcaaaaaca acctcttgtc tcctcttctg accgggacag agtcaccatt tgatttggaa 1680
cagcagtggc aagatctcat gtccatcatg gaaatgcagg aaggtgcggt tggctacagc 1740
tctgactctg agaccctgga tctggaagag gccgagggtg ctgtgggcta ccagcctgag 1800
tattccaagt tctgccgcat gagctaccag gatccagctc agctctcatg cctgccctac 1860
ctggagcacg tgggccacaa ccacacatac aacatggcac ccagtgccct ggactcagcc 1920
gacctgccac caccagtgcc cctcaagaaa ggcagcaagg agaagcaggc tgacttcctg 1980
gacaagcaga tgagccggga tgagcaccga gcccagacca tgaagatccc tttcaccaat 2040
gacaaaatca tcaacctgcc tgtggaggag ttcaatgaac tgctgtccaa ataccagttg 2100
agtgaagccc agctgagcct catccgagac atccggcgcc ggggcaagaa caagatggcg 2160
gcgcagaact gccgcaagcg caagctggac accatcctga atctggagcg tgatgtggag 2220
gacctgcagc gtgacaaagc ccggtgctg cgaggagaaag tggagtccct gcgctccctg 2280
cgacagatga agcagaaggt ccagagcctg taccaggagg tgtttgggcg gctgcgagat 2340
gagaacggac gacctactc gccagtcag tatgcgtcc agtacgccg ggacggcagt 2400
gtcctcctca tccccgcac gatggccgac cagcaggccc ggcggcagga gaggaagcca 2460
aaggaccgga gaaagtgagc ctggggaaga agggggtttg aagcccacca agaccgaaac 2520
tggagaaggy ctgggacctg gacctggacc tggacctaca gcggggactt aaatgccttc 2580
ttatccaata tatcttctca gatgggatga ctgcgggtca gtgtacagga agaggcaggc 2640
actggctggc tcagctccac tcgggtggag tggaaagtggc cagaccattt agacggacag 2700
ggctctcacc ctacccttt cctgtgaggc aggggtggtg gtggagtgtc tggaggtaga 2760
ggagctatgt ggagcaaagg ccgacagagg ggaaggaatg gacctgtgag aggaagggaa 2820
ggtggcagaa agtctcattt cag 2843

```

&lt;210&gt; 70

&lt;211&gt; 2482

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 4176657CB1

&lt;400&gt; 70

```

gtgacgcttt ccctgggctc tgaaggccag ggggcttggga tgtgagtcta gtgctcctgg 60
tccccaaagt cacaaccctt tctatcccag catcccagac ttctccgggc ggggacagag 120
ttggacgtgc atggctgtgt cctgcaatcc cctcgccagc accgcccggg acggagtggga 180
agcgggtccg tgcccccggt cctcgcgcca gcctcgcgcg tggcaagcgc tgtccccag 240
cagaggtgtc cctggatata ggatagcaaa acctgatctc ataaaaacct ggtcacaaaag 300
gacagccctg caaaacagac cctatttggga tcaagtgagc cagttcctgg aacctgaata 360
atgactcctg aatcaaggga tactacagat ttgtctccag ggggtaccca ggagatggaa 420
ggcatcgtga tagtgaaggt ggaggaggaa gatgaagaag accattttca aaaggaaaga 480
aacaagtag agtcatcgcc acaagttctc agtcgctcta caactatgaa tgagagagcc 540
ttattgtcat cgtatttagt tgcatataga gtggcaaaag agaaaatggc tcacacagcg 600
gctgaaaaaa ttatccttcc agcatgtatg gacatggtac ggacaatttt tgatgacaaa 660
tcagctgata aactaagaac tatacctctt agtgataata caatatctcg tcgaatctgt 720
acgattgcaa aacatttggga agcaatgctt attacacggc tgcagtccgg tatagacttt 780
gcaatccaac tcgatgagag cactgatatt gcaagttgtc ccacactctt ggtttatgtc 840
agatattgtg ggcaagatga tttgttagag gatctcttat gttgtttaa tttaaattca 900
catataactg aattagattt atttactgaa ttagaaaact gccttcttgg tcagtataaa 960
ttaaactgga aacattgtaa aggaatttca agtgatggaa cagcaaatat gaccggaaaa 1020
cacagcagac ttactgaaaa attgttagaa gcaaccaca acaatgctgt ttggaatcac 1080
tgttttattc atcgagaagc tttggtatcc aaagaaattt caccaagtct gatggatgta 1140
ttgaaaaaat cagtgaaaac tgttaatttt attaaaggaa gctcactgaa tagccgactt 1200

```

```

ctcgaaatat tttgttcaga gattggagtg aaccacaccc acttattggt tcatacagaa 1260
gttcgttggc tttctcaagg aaaagtattg agcagagtat atgaactcag gaacgagatt 1320
tacatttttc tcgttgaaaa gcaatctcat ttggcacaata tttttgaaga cgacatttgg 1380
gtaacaaaat tggcatattt aagtgatatt tttggcattc ttaatgaatt aagcctgaaa 1440
atgcagggga aaaacaatga tatatttcag tatcttgaac atattctagg attccaaaag 1500
acgttattat tgtggcaagc aagacttaaa agtaaccgcc ctagctacta tatgtttcca 1560
acattattgc aacacatcga agagaacatt attaatgaag actgcttaaa agaaataaaa 1620
ttagagatat tgttgcattc cacttctttg tctcaaaact ttaattatta ctttccggaa 1680
gagaaatttg aatcattaaa ggaaaatatt tggatgaaag atccatttgc ttttcaaaac 1740
ccagaatcaa taattgagtt aaacttggag cctgaagaag agaataaatt attgcagctc 1800
agttcatcat tcacactaaa gaattattat aagatattaa gtttatcagc attttggatt 1860
aagattaaag atgactttcc actgctaagt aggaagagta tattgctggt actaccattc 1920
acaactacat atttgtgtga actaggattt tcaatcttga cacgggttaa aacaaagaag 1980
agaaataggc tcaatagtgc accagatatg cgggtagcat tatcttcatg tgttcctgac 2040
tggaaggaaac ttatgaacag acaagcacac ccatcacatt aaatacaaac tttacaaaat 2100
tctgtgtata gccaggtgtg gtggcttacg cctgtaatcc cagcagtggtg agaccgaggt 2160
gggcagatca cttgagttca agaccagcct ggccaacatg gtgaaacccc atctctacta 2220
aaaatagaaa ccttagccag gcgtggtggc acatgcctgc agtcccagtt acttgggtgc 2280
ctgaggcagc agaattctct aaaccaggaa ggcagagatt gcagtgagct gagataatcc 2340
cactgcattc cagcctgggc aacagcgtga gacttcatct caaaaaaaaa aaattgtatt 2400
tgtactttta aagggatttt gcagtatgtt gtagttaaac gttaataaaa ttatatttgt 2460
aattaggaaa aaaaaaaaaa aa 2482

```

&lt;210&gt; 71

&lt;211&gt; 1730

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7506056CB1

&lt;400&gt; 71

```

gtcccccttg ttcgtcttcg cttttctttc tacttattct tatctgtgtc tttcgctttg 60
tttgcccttc cgtctgtttt ccctcagggc ccccttcttt cctcgacctt ttcaaactcg 120
aaatatggcg ccggagcggc tgcggagccg ggcgctctcc gccttcaagt tgcggggcct 180
gctgctccgt ggtgaagcta ttaagtacct cacagaagct cttcagtcta tcagtgaatt 240
agagcttgaa gataaactgg aaaagataat taatgcagtt gagaagcaac cttgtgtatc 300
aaacatgatt gaacgatctg tgggtgaagc agcagtcagc gaatgcagtc agtctgttga 360
tgaaactatt ctgttaatga ccaaccaccc tgcaccaaatt ttatttggaa caccaagaga 420
taaagcagag atgtttcgtg agcgatatac ctttttgac cagaggacct acaggcatga 480
attatttact cctccggtga taggttctca ccctgatgaa agcggaagca aattccagct 540
taaaacaata gaaaccttat tgggtagtac aacaaaaatc ggagatgcga ttgttcttgg 600
aatgataacg cagttaaaag agggaaaatt ttttctggaa gatcctactg gaacagtcca 660
actagacctt agtaaaagtc agttccatag tgggtttatac acagaggcat gctttgtctt 720
agcagaaggt tggtttgaag atcaagtgtt tcatgtcaat gcctttggat ttccaccac 780
tgagccctct agtactacta gggcatacta tggaaatatt aatttttttg gaggtccttc 840
taatacatct gtgaagactt ctgcaaaact aaaacagcta gaagaggaga ataaagatgc 900
tatgtttgtg tttttatctg atgtttgggt ggaccaggtg gaagtattgg aaaaacttcg 960
cataatgttt gctggttatt caccagcacc tccaacctgc tttattctgt gtggtaat 1020
ttcatctgca ccatatggaa aaaatcaagt tcaagctttg aaagattccc taaaaacttt 1080
ggcagatata atatgtgaat accagatat tcaacaaagt agtcgttttg tgtttgtacc 1140
tggtccagag gatcctggat ttggttccat cttaccaagg ccaccacttg ctgaaagcat 1200
cactaatgaa ttcagacaaa gggtaaccatt ttcagttttt actactaatc cttgcagaat 1260
tcagtactgt acacaggaaa ttactgtctt ccgtgaagac ttagtaataa aaatgtgcag 1320
aaactgcgtc cgttttctta gcagcaattt ggctattcct aatcactttg taaagactat 1380

```



```

cttatcccaa ggacatctga ctccctacc tctttatgtc tgcccagtggt attgggcata 1440
tgactatgct ttgagagtgt atcctgtgcc cgatctactt gtcattgcag acaaatatga 1500
tccttttact acgacaaata ccgaatgcct ctgcataaac cctgggtcctt ttccaagaag 1560
tggattttca ttcaaagttt tttatccttc taataagaca gtagaagata gcaaaacttca 1620
aggcttttga gattcttaaa gatcatctga agaaaattca tcagttttct gcttaactct 1680
atatcttatg tgattctgat attacaataa aattatggta aactttaaaa 1730

```

&lt;210&gt; 72

&lt;211&gt; 3119

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7506185CB1

&lt;400&gt; 72

```

cccgggggcct ggctttcagc gtagcaattc tgccggcgaa gaaggtgagc gcagtgcgtgt 60
gtggcagcag agctccttag gacgaggagc agcgggacga ggaagggcag actggtgaaa 120
tcgcaaactg gggtctgtgt ccggcgccgg acccctatct gcaaagggtcc atgggctaagt 180
aaacacagaa ggttggtgcc atccattttc cttttccctt cacaccctat tccatccagg 240
aagacttcat ggcagagctg taccgggttt tggaggctgg caagattggg atatttgaga 300
gtccaactgg cactgggaag tccttaagtc ttatttgtgg ggccctctct tggctccgtg 360
actttgaaca gaagaagcgt gaagaagagg cacgactcct tgaaactgga actggcccct 420
tacatgatga gaaagatgaa tccctgtgtc tgtcttcttc ctgcgaaggg gctgcaggca 480
ccccgaggcc tgctggagaa ccggcctggg ttactcagtt tgtgcagaag aaagaagaga 540
gggacctggt ggaccgacta aaggcggagc aggccaggag gaagcagcga gaagaacgcc 600
tgcagcagct gcagcacagg gtgcagctca agtatgcagc caagcgctg aggcaggag 660
aagaagaaa agagaatctc ctccgcctca gcaggagat gctagagaca ggcccggagg 720
ctgagcggct ggagcagctg gagtctgggg aggaggagct ggtcctcgcc gaatacagga 780
gtgatgagga gaaaaagggt gcgagcagag tggatgagga tgaggatgac ctggagggaag 840
aacacataac taagatttat tactgtagtc ggacacactc ccagctggcc cagtttgtgc 900
atgaggtgaa gaagagcccc tttggcaagg atgttcggct ggtctccctt ggctcccgcc 960
agaacctttg tgtaaatgaa gacgtgaaaa gcctagggttc tgtgcagctt atcaacgacc 1020
gctgtgtgga catgcagaga agcaggcacg agaagaagaa aggagctgag gaggagaagc 1080
caaagaggag gaggcaggag aagcaggcag cctgcccctt ctacaaccac gaggcagatgg 1140
gccttctccg ggatgaggcc ctggcagagg tgaaggacat ggagcagctg ctggcccctt 1200
ggaaggaggc ccgggcctgt ccctattacg ggagccgcct tgccatccct gcagcccagc 1260
tgggtggtgt gccctatcag atgtgtgtgc atgcccgcac tcggcaggcc gcgggcatcc 1320
ggctgcagga ccagggtgtg atcatcgacg aggcgcacaa cctgatcgac accatcacgg 1380
gcatgcacag cgtggaggtc agcggctccc agctctgcca ggccattccc cagctgctgc 1440
agtacgtgga gcgatacggg aagcgtttga aggccaaaga cctgatgtac ctgaagcaga 1500
tcctgtatct gctggagaaa ttcgtggctg tgctaggggg gaacattaag caaaatccca 1560
atacacagag tctgtcacag acagggacgg agctgaagac catcaacgac tttctcttcc 1620
agagccagat cgacaacatc aacctgttca aggtgcagcg atactgtgag aagagcatga 1680
tcagcagaaa gctctttgga ttactgaac ggtacggagc agtgttctca tcccgggagc 1740
agcccaaact ggctgggttt cagcaattcc tgcagagcct gcagcccagg acgactgaag 1800
ctcttgagc ccctgcagac gagagtcagg ccagcacctt gcgaccagct tctccactga 1860
tgcacatcca aggtctctct gcagctctca ctacggccaa ccaggacggc aggggtcatcc 1920
tgagccgcca aggcagcctc agtcagagca ccctgaagtt tttgtcctg aatccagctg 1980
tgcaatttgc ccaagtgtgt aaggaatgcc gggcagtggt cattgcgggg ggtaccatgc 2040
agccgggtgt tgacttcgg cagcagctgc tggcctgtgc cggggtggaa gctgagcgcg 2100
tgggtggagt ttcctgtggt cacgtgatcc ctccagacaa catcctgccc ctgctcatct 2160
gcagcgggat ctccaaccag ccgctggaat tcacgttcca gaaaagagag ctgcctcaga 2220
tgatggacga ggtgggtcgc attctctgta acctgtgcgg tgtggttctt ggaggggtgg 2280
tctgtttctt cccctcctac gagtacctgc gccagggtcca tgcccactgg gagaaggggt 2340

```

```

gcctgctggg ccgtctggct gccaggaaga agatattcca ggaacctaag agcgcacacc 2400
aggtggagca ggtgctgctg gcatattcca ggtgcatcca ggctgtggc caggagagag 2460
gccaggtgac aggggccctg ctctctctct tgggtggagg aaagatgagt gaagggatca 2520
acttctctga caacctaggc cgtctgcag agctgcagga gaagatggc tacttgatc 2580
aaacctccc cagagcccc ggccaggcac cccagggaag ggctctggg gagaacctgt 2640
gcatgaaggc cgtcaaccag tccataggca gggccatcag gcaccagaag gatattgcca 2700
gcgtagtgt cctggaccag cgatatgcc ggccccctgt cctggccaag ctgccggcct 2760
ggatccgagc ccgtgtggag gtcaaagcta cctttggccc cgccattgct gctgtgcaga 2820
agtttcaccg ggagaagtcg gcctcttcct gatgggcaac cacaccactg cctggcgccg 2880
tgcccttcct ttgtctgcc cgctggagac agtgtttgtc gtggcgctgg tctgcgggga 2940
tcctgttaca aagggtgaaac ccaggaggag agtgtggagt ccagagtgtc gccaggaccc 3000
aggcacaggc gttagctccc gtaggagaaa atgggggaat cctgaatgaa cagtgggtcc 3060
tggtgtcctc tggggcattc cagggcagct cccctcctgg aatagaatct ttctttcca 3119

```

&lt;210&gt; 73

&lt;211&gt; 1881

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 8096611CB1

&lt;400&gt; 73

```

ggggcaggtc ccggaccgca caagctgact tggcgactcc tcgcagcccc tgctctggca 60
cgccacacct cagacgaggc tttgaaattt gttaaataca agtaaggaac acaggcaaca 120
tgtttacac acatgtgctt atgagtaaac gagggccatt ggccaaaata tggcttgca 180
ctcactggga gaagaaactc acaaaggccc atgtatttga atgtaatcta gagataacca 240
ttgaaaaaat tctttcacc caggtgaaaa tagcacttcg aacttcagga caccttcttt 300
tgggagttgt tcgaatctat aacagggaagg caaaatatct tttggcagat tgcagtgaag 360
catttcttaa aatgaagatg acatttcgcc caggactggt tgaccttcca aaagagaatt 420
ttgaagcatc ttacaatgct atcacattgc cagaagaatt tcatgatttt gacacccaaa 480
atatgaatgc tattgatgtt tcagaacact ttactcagaa ccaaagcaga ccagaagaaa 540
tcactcttag agaaaatttg gacaatgatc tacttttcca agctgagagc tttggggagg 600
aatctgaaat tctcagaaga catagcttct ttgatgacaa catattactg aattccagt 660
gtcctttaat tgaacatagt tctggaagcc tctactggaga acgatctcta ttctatgaca 720
gtggagatgg gtttggagat gaaggagctg caggagaaat gattgacaat ctattgcaag 780
atgatcagaa tatcctgtta gaagacatgc atttgaacag agaaatttcc ctgccttctg 840
agcctcccaa tagtttagca gttgaaccag ataactcaga gtgtatatgt gtacctgaaa 900
atgaaaaaat gaatgaaca atattattat caactgaaga ggaaggattt acccttgatc 960
caattgatat ttcagacatt gctgagaaaa ggaaaggcaa aaagaggaga ttgctcatag 1020
atcctatcaa ggagctcagt agcaaagtta tacataaaca gcttacttcc tttgcggaca 1080
cactcatggt tttggaactt gcacctccta ccaaagatt gatgatgtgg aagaagaggg 1140
gaggagtgc taaacttctg tcaactgctg cccaggattt gattcatgct gaactgaaaa 1200
tgttgtttac aaaatgcttt ctgtcctctg gctttaaact tggaagaaaa atgatacaga 1260
aggagtcagt aagggaagaa gtgggaaacc aaaatatagt agagacatcc atgatgcaag 1320
agccaaatta ccagcaagag ttaagtaaac ccaaacttg gaaggatgtg attgggtggat 1380
ctcagcatag ctctcatgag gataccaata aaaatattaa ctctgaggat attgttgaaa 1440
tgggtgtctt agctgctgag gaatcatctt taatgaatga cttatttgca caagaaattg 1500
aatatagtcc agttgaattg gagtcatgta gcaatgaaga aaatattgag acagagagat 1560
ggaatggaag aatacttcag atgttaaata gtttacggga atctaacaag atgggaatgc 1620
agtcccttag tctgtgaag ctctgtagaa atagtgaccg aaaacaagca gctgccaaat 1680
tttatagctt tctgtccta aagaacacgc tggctattga gctgagccag agtgctccct 1740
atgcagatat tatagctacg atgggaccaa tgttttataa catatgaagg aaaccagac 1800
atacagattt atggcatcac tggaaattct gtgtagattg ttcaatttaa tgcagaagcc 1860
atctggaagc agctgaaggc c

```

<210> 74  
<211> 1970  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 8174603CB1

<400> 74  
gcggccatgg cttcggcttc tcagggcgct gacgacgacg gcagtcgtag gaagccccgc 60  
ctggctgcat cgttcgagat cagccccag ccccgcccct ggcgaccgct acccgcccag 120  
gccccaaagt cctgggagcc ctgcgcccgc tgcccgcgac gtatcgctgc ggagggcggg 180  
tgggcccagg tctcgggtgt gagggacagt gggcctgggg caggggcccg agtcggggaa 240  
ctggggggcag cccgggcggt ggaaaacttg ggggaacaga tggggaaggc cccgagggtc 300  
cctgtgcccc cagcaggggt cagcctgccg ctcaaagacc cacctgccag ccaggccgtg 360  
tccttgctca cggagtacgc ggccagcctg ggcctcttcc tgctcttccg ggaggaccag 420  
ccaccagggt aggccgggtc ctgcttcccc ttctcgtga gcgcggaact ggatgggggtg 480  
gtctgccctg cgggcactgc gaatagcaag acggaggcca aacagcaggc aacgctctct 540  
gccctctgct acatccggag tcagctggag aaccagtggt gcccccttct acctgcagtc 600  
tctcgccctg gcgcagagaa catcctgacc catgagcagc gctgcgcagc gttggtgagc 660  
gccggccttg acctcctgtt ggacgagcgc tcgccatact gggcctgtaa ggggactgtg 720  
gctggagtca tcctggagag ggagatcccg cgtgccaggg gccacgtgaa ggagatctac 780  
aagctggttg ctctggggac cggcagcagc tgctgtgctg gctggctgga gttctcgggc 840  
cagcagctcc acgactgcca tggcctgggt atcgcccga gggccctgct gaggttcttg 900  
ttccggcagc tcctgctggc cacacagggg ggcccgaagg gcaaggagca gtccgtgctg 960  
gccccccagc cagggcccgg accccattc accctcaagc cccgcgtctt cctgcacctc 1020  
tacatcagca acacccccaa gggcgcggcc cgtgacatct acctgcccc cacctcggaa 1080  
ggtggcctcc cgcacagccc acccatgcgc ctgcaggccc atgtgctcgg gcagctgaag 1140  
cctgtgtgct acgtggcgcc ctcgctctgt gacaccacg tgggctgcct gtcagccagc 1200  
gacaagctgg cacgctgggc cgtgctgggg ctgggtggtg cctgctggc ccacctggtg 1260  
tccccactct acagcaccag cctcatcctg gctgactcat gccacgacct tccgactctg 1320  
agcagggcca tccacaccg gccctgcctg gacagtgtcc tggggccatg cctgccacct 1380  
ccctacgtcc ggaccgcctt gcacctgtt gcagggcccc cgggtggccc ttccgaacct 1440  
accctgaca cctgcccgtg cctgagcctc aactggagcc tgggggacct tggcatcgag 1500  
gttgtggatg tggccaccgg gcgtgtgaag gccaatgccg cctgggggcc tccctcccgt 1560  
ctctgcaagg cctcctttct ccgggccttt caccaggcgg ccagggtctg ggggaagccc 1620  
tacctcctgg ccttgaagac ctacagggt gccaaaggct ggcctacca ggaggctcgc 1680  
cggcagctgt ctctctcct ggaccagcag ggcctggggg cttggccctc gaagccactg 1740  
gtgggcaaat tcagaaactg aagccagcct cggcgggacc gaggtcccgg agccaagctg 1800  
taccctgct gggggagtgc cctatctgag gacgctgtg gggagaacat gggttgtgct 1860  
gggtgaaact ggggctggag ggaaggagga gcctgctgtg gttgggaggc ggctgctgca 1920  
cgtttgggct tgaataaaga agtttttctg gttaaaaaaa aaaaaaaaaa 1970

<210> 75  
<211> 2968  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 3101042CB1

<400> 75  
gtatcagcac acacctagta tagattaatc tgtgcctatt tctggccata ttttataatt 60  
tgattaatac tgtccttttg gaatccctga tccaatagga ttagtgctct tatgagagcc 120

```

tcaacagagc ttgctctcat tctccctctg ccacgtaagg acacagtga agcagagctg 180
tctgcaagcc aagagaagaa acctgaatga aacctatctt ggtgggtactt gatccttagac 240
tttccaacct ccaggacagg aagaaataag tttatcttgt ttaatctaag cagactgtgg 300
tgttttatta aggcagccca aacagactaa gaccagtcac ttgtgtggat gattgacgtg 360
gcatgtgggt gatgttttag gaaaggggtga cattcaagga tgtggctgtt gtcttcacca 420
aggaagagct ggcactattg gataaagccc agataaacct gtaccaagat gtgatgctgg 480
aaaacttcag gaacctcatg ttagtgagag acgggattaa aaacaacatt ttgaatcttc 540
aggcaaaagg gttaagttac ctttcgcaag aagtgcctca ttgctggcag atttggaac 600
aaaggatccg ggatttaact gtgagtcagg attatatcgt gaaccttcaa gaagagtgtt 660
ccccacattt agaagatgtt tccctcagtg aagagtgggc aggcatttct cttcagattt 720
ctgaaaatga aaactatgta gtaaattgcca ttatcaaaaa tcaagatata acagcatggc 780
aaagcctgac acaggttctt accccagaat cgtggaggaa agccaacata atgaccgagc 840
cccagaactc tcaggggaaga tataagggaa tttacatgga agagaaattg tacagacgtg 900
ctcagcatga tgacagcctc agttggacct catgtgatca tcatgagtc caagaatgta 960
aaggagagga ccctggtaga catcccagct gtgggaaaaa cttgggtatg aaatcaacgg 1020
ttgaaaaacg taatgcggcc catgtattac cacagccttt cccatgtaat aactgtgggg 1080
tggcctttgc agatgataca gatcctcatg tccatcacag cactcaccta ggagaaaaat 1140
cttataaatg tgaccagtat ggaagaact ttagtcagag ccaagatctt atcgttcatt 1200
gtaaaactca ctctggcaag actccctatg aattccacga atggcctatg ggctgcaaac 1260
agagctcaga ccttcccaga tatcagaaag tctcctcagg agacaaaccc taaaaatgta 1320
aagaatgtgg caagggcttc aggcgcagct cttcccttca caaccatcat cgagtccaca 1380
caggggagat gccctacaaa tgcgatgaat gtgggaaagg gtttggattt aggtcacttc 1440
tttgatttca tcagggagta cacacaggga aaaagcccta taaatgtgaa gagtgtggga 1500
agggctttga tcagagctcc aaccttcttg tccatcagag agtccacact ggagagaagc 1560
cctacaaatg cagtgaagtgt ggcaagtgtt ttagttcaag ctccgttctt caagtccact 1620
ggaggtttca cacaggggag aaaccatata ggtgtgtgta gtgtggaag ggcttcagcc 1680
aatgtacaca ccttcacatt caccagagag tccacacagg ggagaaacca taaaaatgta 1740
atgtgtgtgg aaaggatttt gcgtatagct ctgttcttca cactcatcag agagttcaca 1800
ctggagaaaa accatataaa tgtgaagtgt gtggaaagtg cttcagttac agttcatatt 1860
ttcacttaca tcaaagagat cacatcagag agaaaccata taaatgtgat gagtgtggta 1920
aaggcttcag cgggaattca gatcttaatg ttcacctcag agtccacaca agagagaggc 1980
cctataagtg taaggcatgt ggtaagggct tcagtcgtaa ttcataacct cttgcccac 2040
agagagtgca tatagatgag acacagtaca cacactgtga gcgtggaaag gaccttctga 2100
ctcatcaaag actacatgag cagagagaaa cattataaat gtagtaagtc agggttcaat 2160
tagtaaaaaca gaagccacac tgtattccag atgatagagg cttactcagc ctgtgggagg 2220
gctaggggag caaaagacag ggacactgcc atcgataatg tcagcctgca gcaactgaag 2280
gtgccaggca aggactgtgg atttcaagaa cttctgtgaa gctcccatca actgctgtat 2340
gctacaatgg caaagagggt acttttgaca aatgggtggt ttgtggtagt ggttgtagt 2400
agtgtgcaaa tttccattaa aggtgtgtgc caggaaggaa attctaattg ggatgatcat 2460
ggtaagggtc tcagttcagc caacgtcttt agatttttt ttagtcccca gtagaaaaat 2520
actctctgta tgaagaacat tcaaaagtgt gctcaaaaat gaaacccatg ctcttactat 2580
gaaagaacgt tagtaccag gttttccatg agattctcta cacaggcaag aagctccata 2640
gaagtggcat ttgaagggtg tggcagaggc agtgctgtgt ttatcacact ggttccattt 2700
ccttgcaaat aagaagtcta tttcccagta acccttgtag ttaagagtgt gcccatgtga 2760
ttgagttcta gccaatggag tgtgagcaaa agtgatataa gccactttca ggtctagcct 2820
ttacaaacat cctcaggctt ctctatccct gccaaaggtga ccttggaggc tgcttattcc 2880
acgactgggt tgatagaagg tcaactcttc atctgtgttg gatttctttt tagagtaaga 2940
aataaaagtt cattgtgtct caaaaaaa 2968

```

&lt;210&gt; 76

&lt;211&gt; 4497

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 4972035CB1

&lt;400&gt; 76

```

cctcacgtac cacaacctca cgtaacctct cccccccct ccgatacccc cagccccggg 60
atggcagcgc ccgccagcct tcccggccgg accggctgtc gccagtgcgc gctttagcct 120
gtgggactag gccccgccga ggcctgacct ccggagccgg gacctactgt gcagccagca 180
gccgggcccgg gactgagggg ccgctcccc ctcgcgccc gtcgggagga cgctgtcgtc 240
ccggggactg ggggcccggg ggtggggggg agccaggcgc aggaagtcgg gacctgtatt 300
gggaaaaaag cggggccggg gtcaggagcg cccctcccc tcacctcccc cctcccccg 360
gggttggggg ggccggagca gagagcacc agccggggag gtggatgaat gtgggagaaa 420
atggagacca agacgatcgt gtacgacttg gacacatcag gggggctgat ggagcaaate 480
caagctctgc tggctcccc caagacggac gaggcagaaa agcgcagtcg gaagcctgag 540
aaggagcccc ggagaagcgg cagggccacc aaccacgaca gctgcgatag ctgcaaggaa 600
ggtggagatc tcctgtgctg cgaccactgc ccggctgcct tccacctcca gtgctgtaac 660
cctccactga gtgaagaaat gttgcctcct ggagagtggg tgtgtcaccg gtgcaactgt 720
cgccgaaaga aacgagagca gaaaaaggag ctgggtcatg tcaatggact ggtggacaaa 780
tctggcaaac ggactacatc cccagcagt gacactgact tgttggacag atcgccagc 840
aaaactgaac taaggccat tgcccatgcc cggatcctgg aaaggagagc cagcaggcct 900
ggcacacca catccagcgc cagcacagag actccacct ctgagcagaa tgatgtcgac 960
gaagacatca ttgacgtgga tgaggaacca gtagcagcgg agccagacta tgtgcagccc 1020
cagctgaggg ggccctttga gctgctgatt gctgccgcca tggagcggaa cccacccaa 1080
tttcagttgc ccaatgaact gacttgatc actgcactac caggttctag caagaggaga 1140
agaaaggagg aaaccacagg gaaaaatgtt aagaagacac agcatgaatt agatcacaat 1200
ggtctcgttc ccttaccctg caaagtctgc ttcacgtgta acaggagttg ccgtgtggct 1260
cctctcatcc agtgtgacta ttgccctctc ctgtttcaca tggattgcct cgagccgccc 1320
ctcactgcca tgcccctggg cagatggatg tgtccgaatc acatcgaaca tgtggtgctg 1380
aaccagaaga atatgacact gagcaatcgg tgccagggtg ttgatcgttt ccaggacacc 1440
gtttcgcagc atgtcgtcaa agtggacttc ctgaaccgaa tccacaagaa gcacccccct 1500
aaccggcgtg tgctccagtc ggtcaaaaga agaagcttga aggttcctga tgctataaaa 1560
tctcagtacc agtttccacc ccctctcatt gcaccgcgg ccattcggga cggggagctg 1620
atctgcaatg ggatccctga ggaatcacag atgcacctt tgaactctga gcacttagcc 1680
accaagcag agcagcaaga gtggctctgt agtgtgtgtg cgctccagtg cagcatattg 1740
aaacatttat ctgctaagca gatgccttcg cattgggact ctgaacagac agagaaggct 1800
gatattaagc ctgttattgt gactgacagc tcagtcacca cctccctgca aacagctgac 1860
aagacaccta caccttccca ctacccttg tcctgcccct cagggttagg caccagaaat 1920
tccttgagct gctctccacc ccaccagtc ccagccctag aggacatcgg ctgcagttct 1980
tgtgcggaag aatccaagaa aacccttgtt gggactgcca atgggccagt gaacacagag 2040
gtgaaagcta atggcccaca cctctacagc agccctactg attccacgga cccccggcga 2100
cttcaggcgc ctaacacccc actaccaggc ctctcacacc ggcaaggctg gccccggccc 2160
ctcacgccac cagcggctgg gggccttcag aaccacaccg tcggcatcat tgtgaagaca 2220
gagaatgcca ctggcccag ctcttgcccc cagaggagtt tggttcctgt cccaagcctg 2280
cccccttcca tcccagctc ttgtgccagc atcgagaaca ccagcacttt gcaaagaaag 2340
actgtccaat cacagatagg acctccgttg acagattcaa ggccactggg ctacccccca 2400
aacgccaccc ggtgtctcac tccccgcaa gcagcaggag atggtatctt ggccacaaca 2460
gccaaccaac gattcagctc accagcgcga tcgtcagatg gcaaggtcag ccccggcacg 2520
ttatccatag gaagcgcttt aaccgtacct tctttcccag ccaactctac tgccatgggtg 2580
gacctcacca actcacttcg agcatttatg gatgtcaatg gagaaatcga gataaatatg 2640
ctggacgaga agctgatcaa gtttctggcc ttgcagagaa tacatcagct tttcccctcc 2700
cgggtccaac cttcaccggg cagtgtcggg acacatcagc tggcttcttg agggcaccac 2760
atagaagtgc aaagaaagga ggtacaggcc cgagctgtgt tctaccccc cttaggggtg 2820
ggaggagctg tgaacatgtg ctatcgaacc ctctacatcg ggacaggagc tgacatggat 2880
tgtgtcctta caaactatgg tcaactgaac tacgtgtccg ggaacatgc ctgcatattc 2940
tacgatgaga ataccaaaca ttatgagctg ttaaactaca gtgagcatgg gacaacgggtg 3000
gacaatgtgc tgtattcatg tgacttctcg gagaagaccc cgccaacccc cccaagcagt 3060
attgttgcca aagtgcagag tgtcatcagg cggcgccggc accagaagaa ggacgaagag 3120
ccaagtgagg aggcagccat gatgagttcc caggcccgag ggccgcagcg gagaccctgc 3180

```

aattgcaaag ccagcagctc gagcttgatt gggggcagtg gggccggctg ggagggcaca 3240  
gccttactgc accatggcag ctacatcaag ctgggctgcc tgcagtttgt cttcagcatc 3300  
actgagtttg cgaccaaaca gcccacaggc gatgccagcc tgcctgcagga tgggggtcttg 3360  
gccgagaagc tctctctcaa gcccaccag gggcctgtgc tgcgctccaa ctctgttcct 3420  
taggactggc ggctaccccg ccactggcct gtacacccac ccaagactcc tgcaatgcaa 3480  
aaatgtacac aaaccaagcc cgggtgtttt ctatactcta ccagaaaccc ttcaactaca 3540  
atctttgcat gaaatgaaga aaaccttttg actgtttttt aaaaatcctt tttcttttct 3600  
caagttctag ggggcatttg cacatatatt tgtactcaac atttcatggg aaagcggcag 3660  
acctgagctg aggaacagcg tgggcaggga gggaaagacc cagggtctgg acacttcctc 3720  
caacacaaaa cccttcccca cccacctcct gctccctccc cctcgccac cattgtaaaa 3780  
taatcagaaa cttgttctat tttgtggcag tgacaatagt tttatattaa aagaaaaaat 3840  
acagttttca tacagcaaaa tctatacaat atcattgttt tatttaatat aaagatcgct 3900  
accacacctc cttccatggg tcccaccctc cacgttattt tccctttctg cagcgggtgc 3960  
actacaggta gctactgtgt attatggaca aatgagaaat gaattccttt tctggctgtc 4020  
catctatttt atttcaaata aggaaaagtg tatttggatt ttgtgtaaat acatctagt 4080  
atgacatttt ttcaatgttt ttaaaaaccg tgtacagtac tacatgtggg agagcgtttt 4140  
ctcaaattgt ctattgtagc aaaaatgttt ttgtcgtaaa cctgttttgt ctctttttt 4200  
tgttctcttg ccacttctct cctcttctc ctgcccctgg ttccctctc tctcccacc 4260  
cccacaacca gtaccaatgt acatagtaat tgtaatgttt tagactttac agaaactttc 4320  
ctgtattctg tctgtacttt aattacagat agatataata gcatatgctt caaaaaataag 4380  
agcggatttc caccttataa aaggaacatt ttcataatag ctgtattttt gaaagaaatt 4440  
tcacactgtg gcacaaacaa gttctttttt catccccggg tgtggattcc cccaggg 4497

<210> 77

<211> 513

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 7506265CB1

<400> 77

cccacgcgtc cgcccacgcg tccgcaggag gcctacacgc cgccgcttgt gctgcagcca 60  
tgtctctagt gatccctgaa aagttccagc atattttgcg agtactcaac accaacatcg 120  
atgggcgggc gaaaatagcc tttgccatca ctgccattaa ggtggaacgt gtgatcacca 180  
ttatgcagaa tccacgccag tacaagatcc cagactgggt cttgaacaga cagaaggatg 240  
taaaggatgg aaaatacagc caggtcctag ccaatggtct ggacaacaag ctccgtgaag 300  
acctggagcg actgaagaag attcggggcc atagagggtc gcgtcacttc tggggccttc 360  
gtgtccgagg ccagcacacc aagaccactg gccgcggtg ccgcaccgtg ggtgtgtcca 420  
agaagaaata agtctgtagg ccttgtctgt taataaatag tttatatacc aaaaaaaaaa 480  
aaaaagggcg gccgctcgcg atctagaact agt 513

<210> 78

<211> 509

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 7506304CB1

<400> 78

caactcctgt cgtgcccgcc cggagaggaa gcggaaatgc gtgtacggta ttaaaggcgt 60  
cgcctcgccc ctgccttctc ctctttccgt ctcaggtcgc cgctgcgaag ggagccgccg 120  
ccatgtctgc gcatctgcaa tggatggctg tgcggaaact ctccagtttc ctgatcaaga 180

```

ggaataagca gacctacagc actgagccca ataacttgaa ggcccgcaat tccttccgct 240
acaacggact gattcacgcg aagactgcca cctcctatgt gcggaccacc atcaacaaga 300
atgctcgcgc cacgctcagc agcatcagac acatgatccg caagaacaag taccgccccg 360
acctgcgcac ggagcccatc cgcagggcca gcgccatcct gcgcagccag aagcctgtga 420
tggtgaagag gaagcggacc cgccccacca agagctcctg agccccctgc cccagagca 480
ataaagtcag ctggctttct caccaaaaa          509

```

<210> 79

<211> 5881

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 7506198CB1

<400> 79

```

ggagggcagg cggcggtggc agccatgttg ttgtttgatc tgtggatgaa atgaatcatg 60
atthttcaagc tcttgcatga gaatctcggg gaatgggaga gcttttgcct accaaaaagt 120
tttgggaacc tgatgattca acaaaagatg gacaaaaagg catatttctt ggggatgatg 180
aatggagaga gactgcatgg ggagcttctc accattcaat gtcccagcct attatggtac 240
agagaagatc tggacagggt tttcatggaa acagtgaagt aaatgcaata ctgtctccgc 300
gatcagaaaag tggaggcctt ggtgtgagca tggtagaata tgtattaagt tcttctcctg 360
ctgataaaatt ggattctcga tttaggaagg gaaatttttg cactagagat gctgaaacag 420
atggacctga gaaaggagat caaaaaggca aggcttctcc atttgaggag gacaaaaaca 480
gagatcttaa acaaggagat gatgatgatt ctaaaataaa tggcagaggt ttgccaaatg 540
gaatggatgc cgattgcaaa gattttaatc gtactcctgg aagtcgtcaa gcctctccaa 600
ctgaagtagt tgagcgcttg ggccccata ctaatccctc agaaggactg gggcctcttc 660
ctaattcctac agctaataaa ccacttggtg aagaattttc aaatcctgaa actcagaatc 720
tggatgccat ggaacaagtt ggtctggaat ccttacagtt tgactatcct ggtaatcagg 780
taccaatgga ctcttcagga gctactgtag gcctttttga ctacaattcc cagcagcagc 840
tctttcagag gactaatgca ctaacagttc aacagttaac tgcagctcaa cagcagcaat 900
atgcattagc agcagctcag cagccacata tagctggtgt attctcagca ggccttgctc 960
cagctgcatt tgtgccaaat ccatacatga ttagtgctgc tcctccaggg accgatccgt 1020
atactgcagc aggattggct gcagcagcta cattagcagg tccagcagtg gttccacctc 1080
agtattacgg cgttccatgg ggggtgtatc cagccaactt atttcagcag caagctgcag 1140
ctgcggaaca taacacagcc agtcagcaag cagcatcaca agctcagcct ggacagcaac 1200
aggttctccg tgctggagca ggtcagcgtc ctcttactcc caatcagggt cagcaagggc 1260
agcaagcaga atcacttgcg gcagctgcag cagcaaatcc aacattggct tttggtcagg 1320
gtcttgctac tggcatgcca ggctatcaag tactagctcc aactgcctat tatgatcaga 1380
ctggtgcctt agtgggtggc cctggagcaa ggactggcct tggagctcca gttcgggttaa 1440
tggctccaac acctgtttta attagttcag cagcagcaca agctgcagca gcagcagcag 1500
ctggaggaac tgcaagtagc cttacaggca gcacaaatgg tctgtttcgg ccaattggca 1560
ctcagccacc acagcagcag caacagcagc caagcactaa tctgcaatct aattcatttt 1620
atggaagcag ttctttgact aatagctccc agagtgttc tttattttct catggacctg 1680
gtcaacctgg aagtacatct ctgggctttg gaagtggtaa ctctttgggt gctgctatag 1740
gctcagccct cagtggattt ggttcatcag gaggactgac aaatggtagt ggtcgatata 1800
tctctgcagc acctggagca gaagcaaaat atcgaagtgc ttcaagcact tccagtctat 1860
ttagctccag cagccagctc tttcctcctt cccggcttcg gtataatagg tctgatatta 1920
tgcttcttgg ccgcagtaga ttattggaag atttcagaaa caaccgcttc ccaaaccctc 1980
agcttagaga cttgattgga catatagttg agttttctca agaccagcat ggttctagat 2040
tcatacagca aaaactagag agagctactc cagctgagcg acagatggta tttaatgaaa 2100
ttctgcaagc agcctatcaa ttaatgactg atgttttttg caactatgtt atacagaagt 2160
tttttgagtt tgggagctct gatcaaaaat tagccctggc tactcgtatt cgtgggtcatg 2220
ttctaccctt agccttgagc atgtatggct gccgcgttat tcagaaagca ttagaatcta 2280
tttcttctga ccagcaggta attagtgaat tggtaaaagg gctggatggg catgtgctca 2340

```

```

aatgtgtgaa agatcagaat ggaaaccatg ttgtacaaaa atgtatcgaa tgtgttcagc 2400
cacagtcact acagttcatc attgatgctt tcaagggaca agtattttgtg ctttcaactc 2460
atccttatgg ctgcagagta attcagcgca tcctagagca ttgcaactgca gaacagacct 2520
tacctatctt agaagaactc caccaacata cagagcagtt ggtacaggat cagtatggca 2580
attatgttat tcagcatgta ctggaacacg gtcgacctga agacaagagc aaaattgttt 2640
ccgaaatcag gggaaagggt ttagccctga gtcaacacaa atttgccagc aatgtagtag 2700
aaaagtgtgt tactcatgcc tcccgctgct agagagcttt actgattgac gaggtttgct 2760
gccagaatga tggctcctcac agtgccttat acaccatgat gaaggaccag tatgccaat 2820
acgtgggttca aaagatgatt gatatggctg aacctgctca gagaaagata atcatgcaca 2880
agattcgacc tcacattact actttgcgca aatacacata cgggaagcat atactggcca 2940
agttggaaaa gtattatttg aagaatagcc cggacctagg acctattgga ggaccaccaa 3000
atggaatgct gtaaattaca ggagcaagag aaagaagata atttaacct gtgaaaagaa 3060
ttttttttgt gtgtgaatta tcaaaacaca actcaactat gaatcttcaa ttttttttta 3120
aagcaaaaact atttattgac tttattcatc catttgtaaa ttttttaagg ttcttgtgta 3180
tatttggggg gtgggggatg aattataaat tatattcagc cctgagtgga gacctatcag 3240
attggattgc tggcaaagca cagaatgcct gtatatgatg taactgtatc aaaaataaaa 3300
agctgtcaca tattttgtaa atttttacct tgtaaagtca caaaaatagt ttttaaagga 3360
aaaagtacag tattctttta ataaactggc tcacagctcg gtaggtctac aaccccatag 3420
cacaacaggt ttatagagat gtatatagaa ttatagtctt tatttttttc ctttgcgtga 3480
aaccttttat aacagattaa caatcaactg cataaatatt attaataatt taaaaagagt 3540
taagttgtat tttgataatt cacaaactat catgcaaata acgagtaagt agacaagaat 3600
aaagtgggtt gagatgaaaa gaacctaaac ttatttacag tagatgtggg tttaatacaa 3660
ttactgccct aaaaagtctc tggcaatgta cagaaatatt gtatatactt acatatgtaa 3720
ttgttgtaag agttaataac aaaatcatgg tgacacttcc aattaagtgc actaaatgaa 3780
aagttaagtc acttattaac ttttcagttt ggtttgcaat gagaaagagt ggaaatttgt 3840
attttgtttt gcttatagaa ttacagacat gttgaggaag tgttgagctt tattttgctt 3900
tttcatagag gcagaaagta ggaaccagat agagatgaaa agggggccact gaaaagtga 3960
tttgatagct cagcatttaa gcatgattac atattcagat agctcttttt gctttctata 4020
aatatagca ttgtgtgtgt agtaatagat gtaagtttac actttgaaag gaaatctgt 4080
ttcaatgttt attataaaa ccttgctaatt ttagtagtga tgctttcctt ggttgtagag 4140
gtgtacattt gtaaaccctt atgctgtaaa tggaaattgt tttatctctt tgggatacat 4200
ttgcatttta gtgtacattt acgtccctgc cctctttgac ctggcaatat agtggtgtat 4260
aatgtaaatt tatttctcca aatcgagagt gattttttta aaatttttta tctttatatg 4320
gtttcagaag tatgaaccag ctttcttttt attattgtga gatcattttg ttttataaca 4380
tagttgttga ctgttaatat ggacctgcta gaatttggat cactttcaat tgaagtcagg 4440
gtattgtgca taatagaaa tattggactg agatatttgg ttaccatgga ggccaatgct 4500
tttttcatct tattaatagt gatgtgactt ttttctttgt acagaagagt actgtatttt 4560
tgaatagcct actcccaagt aagagcaaat ctgtatgata acattttttc ctctggacat 4620
aagacataac agtaacacga tgtacattta caagcggcct tatgtacatt tcccaacaa 4680
ctttttaagg caaaattgtg accatatgtg tataattaaa atcgttttta atcctttgcc 4740
tatgaaaata ttttggaaaa aaacttgctg tgtatattca gtttctgaaa gataaagaaa 4800
gtgctttgta ttttgttgaa gtcagtattt tgtataaaca tttatgttga ccacttatg 4860
ttcagtgtcg aaaactaaaa tgaacatgct attctgtcag ctgaatatgg aagagatctt 4920
tttttactag agatctgcag aagaaacgca atcttctgag cacaatatgg aatctaaagg 4980
ttttatcact tagttgttca tattatgaac ctaaaaataa tggcataaag tttggggatg 5040
ccaggcatac tttttcatgt ttggtgttga gttattttac ttttctaacc caacattcct 5100
tggtgagacc attaaatcca aacacttgtc accgttcctt ctcatagtea ctctgggtca 5160
tcagcatgtc ccagtcactg cagcaatgcc ttgtgtttgt ttcatttttt taaaaccac 5220
acaaagccgc tgtctcactt tttcctactt taccaacctc agagtatttc ggcccgatc 5280
gaacttttgt tctcagtatc agcccatggg ttccaggatca aagctgtcat gttggagatt 5340
ggtaatggct ttctgtctt tgtacagttg aattcttagt cttccttcat ccttgccctc 5400
tgttggcaca ggcattatct ctgcaatttt agaaaatgac aagtagagaa tactacattg 5460
agaaaactaa cctcttctt ggggtcctga tactcattcc catttgtccc agtgctgaca 5520
acccaatctt cccaatactt tcaggcctgc tctacaaaag tacctgttct tgtagaaatt 5580
ttacagctcg ccattttggg tgcccccccc aatttttacc ttttagtaag ttggcatgaa 5640
attttggtaa aatctgaaaa tcacatttca gaataaaaca attgggcaaa actacctagg 5700

```



ctttactctt gagtgtctcc ttttgatagg gattgtttct ggaccagttt gtctaagtcc 5760  
tggctcttat tggttcatat gaaataatgt taacttcact tctttgtata ttatgtataa 5820  
attagaaaat gaaaaatgtg tgaataacat tgtatgaaat aaacctgggc ttgtgttttt 5880  
c 5881

<210> 80

<211> 5712

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 1381261CB1

<400> 80

cagaaggaag agggccagca gaagggactc gctgagccac cccctgcccc actccatgtt 60  
ggctgtgtctc tcaggctggc accatgcccc cccctgcaga ggtgacggac ccgtcccatg 120  
ccccgcctgt cctgcgccag ctcaatgaac agcggctccg tggcctcttc tgtgacgtca 180  
ccctccatag cggagacacc aagtccctg ctacccgag cgtcctggct gcttcaagtc 240  
ccttcttcag agaggccctg ctacttcag cccactacc ccttccacca gctactgggg 300  
gcgcgcgacc caaccctgcc accaccacag ctgcctcttc ctctctcttc tcttctctgt 360  
cttcttcttc tcttctcttc tctgtctctt cttcttcttc ctcttctctt tcttctcttc 420  
ctccagcttc tccccctgt tcttccccac cccgggtcct ggagttgcca ggagtccag 480  
cagctgcgtt ttctgatgtc ctcaacttca tctacagcgc cgggctcgca ctgcctgggt 540  
gtggagggga cggggcagct gtagcagaga ttggagcttt ggggcgcctg ctgggcatct 600  
ccgccttca gggcctgggg gagggaggtg atgcctgggt acctcctacc ccagcccca 660  
tggccacctc gcagcctgaa gaggacagct ttgggcccgg gccagggcca gctggggagt 720  
gggaggggtga cagggtgag gccagggccc ctgacttgca gtgctccctg ccccgggcgc 780  
ccctccctgt ccccatgtgt ggaaaaagct tcatccatcc caaacggctg cagacccatg 840  
aggcccatgt cccagcaggg gccagcacgc gggggtctac aggggtggga gctggggcgc 900  
ctggccctgg tggctctgca ggggtggagc cctcagccct gcctccacca gtgggcttcc 960  
gagggggccc cgagcacgtg gtgaagggtg tgggcggcca cgtgctgtat gtgtgcgcgg 1020  
cctgcgagcg tctctacgtg acctgttcca gtctgaagag acacagcaat gtactactgt 1080  
ggcggaggaa gtaccctgc cgctattgtg agaaagtgtt tgctctggcg gactaccgca 1140  
cgaagcatga agtgtggcac acgggggagc gcaggtagca gtgcatcttc tgttgggaga 1200  
cctttgtcac ttactataac ctgaagaccc accagcgagc cttccacggc attagcccgg 1260  
gcctccttgc cagtgaagaag acaccaatg gaggctacaa gcccaagctc aataactca 1320  
agctctaccg cctgctcccc atgcgggcag ccaagcggcc ctacaagacc tacagccagg 1380  
gagccccgga ggctccccct tctccaaccc tcaacacacc ggccctgtg gcaatgccag 1440  
ccagcccgcc gcctgggcct ccacctgcc cagagcctgg ccttccaccc tctgtcatca 1500  
cttttgccca cccagcccc tctgtcattg tccatggggg cagtagcagt ggtggagggg 1560  
ggagtgggac ggccagcaca ggagggtccc aagctgcctc gggtatcact tacactgctc 1620  
ccccgaggcc acccaagaaa cgagaatacc cacctcctcc ccctgagcct gcagccacac 1680  
ccaccagccc agccacagca gtcagcccag ccaccgtgc agggccagcc atggccacca 1740  
ccacggagga ggccaagggc cggaatccac gggctggaag gactctgact tacacagcca 1800  
agccagtggg cgggatttgt ggaggtgggg gtccccccac aggggctggc cggggccctt 1860  
ctcagctgca ggctccacct ccactgtgtc agatcactgt gcgaataggg gaggaggcca 1920  
tcgtcaagcg ccgcatctca gagactgacc tgcgtcctgg ggagctgagc ggagaggaga 1980  
tgaggagag tgaggaggac gaagaggagg aggacgaaga ggaggaggag gaggatgagg 2040  
aggaatcaaa ggctgggtggg gaggaccagc tctggaggcc ctactactcc tacaagccta 2100  
agcgcaaggc tggagctgct ggaggtgcca gtgtgggggg cagtgggctg ccccgaggcc 2160  
cccggccacc acgttggagg cagaagctgg aacggaggag ctgggaggaa accccagcgg 2220  
ccgagagccc agcgggacgt gccgcacag agcggaggca ccgatgcggg gactgtgccc 2280  
agaccttcac caccctgaga aagctgcgga agcaccaaga ggccacgggt gggggctccc 2340  
acagctcccc ggccggacgg agggcctcca cccgctttac ctgccccccac tgcgccaagg 2400  
tgtgcaagac cgcagctgcc ctgagccgcc acgggcagag gcatgctgct gagcggcccc 2460

```

ggggcacccc aaccctgtc attgcctatt ccaagggcag cgctggcacc agggccgggg 2520
atgtcaagga ggaagcccc caagagatgc aagtctcttc atccagcggg gaggcagggt 2580
gcgggagcac tgctgctgag gaagcttccg agaccgcctc actccaggac cctatcattt 2640
caggggggtga ggagccccc gtagtggcaa gcgggggcag ctatgtatac ccacctgtgc 2700
aggaatttcc actggccttg attgggggcg gccgggaacc tggcggtggc aggggaaaat 2760
ctgggagtga agggccagtg ggggctggtg agggggaccg gatggagggg ataggggctg 2820
ccaaagtcac tttctaccct gagccctacc cgctcgtcta tggcccccag ctccctgccc 2880
cctaccctta caacttcagt aacttggccg ctctcccgtt tgctctcaac atggctctac 2940
ctgatgagaa ggggtgcgggg gcccttccct tcctaccagg ggtctttggc tacgcagtga 3000
atcctcaagc agcaccacct gcccaccaa caccacctcc cccaactctt cctccaccaa 3060
ttccccctaa gggagaaggg gaaagggcag ggggtgagag aaccagaag ggcatgttg 3120
gggtgaacctt ggggtcfaat ccccttttca ccagatgcca cctccctga acccccacc 3180
actaccagct ccctggcctc cctgcccctt gggagcccct tcacactctt gtgcagggac 3240
ttggggggccc ctggagctca ggggtcaggc tgctttgtgt gagatgtagt ttcccatct 3300
cctgggaagg gatctttcga ggttcccctc tcagtcttcc tccagggaat ggcctccatg 3360
aggggcaggg ccagcttcca tcccttctcc agcccttggg gcaactgagc aatatactta 3420
acctgaatct ctactcacag cccccaccag ctctgaatgt ctaacctgct cccctgattc 3480
gtaaacctag gggaaacctt ctctctcacc taatgaccgg ccttgttctg aagctttctc 3540
taagcccttc ccagtgtctt cctagcacat gccattcttt gtggcccagg gcctggacca 3600
gaccattgtg atacctgacc ccgcccacct tggagtgttg ctttgggttt catccttccc 3660
cagcgtgggt ctctacgtcc ctgtttccct tgtatcaaga caccttctc agcttccatg 3720
cctttggatc ttccatgttc ctccccatat tectggactt cggagatggc ctctcccaag 3780
ccagggtcaag gaggtttggg ggagggttgc ccctctgccc ctctgttctg ttggtgagca 3840
ctttcccatg ccagggcagg gaaatatttg ccctatcttg acccccaaat ccagttagct 3900
ccagattctt ccaaggcaaa agaggtaagc agatcacacc tctttctgcc tctacatatg 3960
gcctattctg ggctagacca gatttggggg ccaggaggga agaactccat atgggatgga 4020
gaagggaatc tacttttctc ctgttttttt ttctgatgg tttctcccag actagaccaa 4080
atagccagaa aaatgatagg ggtcggatgg gtgggtaagc ccaggatttg cacatgacct 4140
tccatcctta cctgtattcc catctcccca gtgtcactcc cctcaccaat cactccagat 4200
ggttttgggg gaaccttct actcttctgg tgggctttgg ggtatcccca ccaactttcc 4260
ttcaaaatg ggaacaaac acaaaaaaaa aatcactaaa aattcccaca aatcttgttt 4320
cctaggatg ggggtacagg actttaaata gtccctaata cctaatttgc actagttaac 4380
cctggtcagg gtccctgtat ttcttccag tgggggagat aaatgtttgc tcctaattct 4440
ctttgaaaac tgggcctccc tgctctgtga ttggataaat atttcccatc ccaccacct 4500
cccccaaaa aatagctcac aaggggagag ccagtatggg ggagcaaatt tgacaaatgg 4560
gaattagagg agtgcagttt taaaaggaaa agttgctgtc atcaaaatgg cagccttttc 4620
cccagctact gtttttgggg ccaagatggc tgccctagca gcaatcactg ccaagggcaa 4680
gatcatggct tttggaggga ggtgagttta gggagggcca ggaccatcct cctaccctc 4740
ataccctccc agcatataca aaaggggagg ttttagacag gctccctgaa tgtaaccac 4800
agaggagtca ctccctcatt cctcctctgt ctctttgcac ttttcttggc cttggccaca 4860
gcctgagtga cgaatttctt actgaatgta ccaagttcca attttaagg ggggaaagg 4920
tttcaaatgg ggaacaaac acaaaaaaaa aatcactaaa aattcccaca aatcttgttt 4980
ctggcacttt agaaaaactg caaaaaaata cgtaataaag aatacatata tatatatcta 5040
cacacaaatt atatatctat ctatctatac agcggaaacca caagagagac tgaggaaggc 5100
ctggaggcag gggcagaggg gacgacagtg cccctatata ctttaacctat actcctctga 5160
ggcaaacagg catgggaaaa tgggaagggtt gaggatggac cggagaattg gaacttcaga 5220
atagggtcaaa attccaaaac catggacatt tttttttggg agaattgaga ttgtagacat 5280
tttttttttc ttaaatatga tcaaggaaaa tagcttccag aatgtggtgg ttctgggcaa 5340
caaatgagat tgtggcgacg tggagattaa aatatatgta tttgagctgg ggaatttgaa 5400
tattgtgagt ttcagatgtt ggaattttgg gattttgcag ttttgcctt tgaaaatgat 5460
caagtcttgt cagttcgtgc cctcttccc catgttccct ggggaagacgg gtggtggcag 5520
agtgaagagg cactgggttc tgtgccgacg cacgcaaaat ttagaattct acagactagc 5580
tctatacgta gtgaggacc agatttagag aaactgacca atatttatct ccgcatttgg 5640
gtgtgtgtcc aactctgtag gccataaac caacaagaca aatgaactgt gctcaaaaaa 5700
aaaaaaaaaa aa 5712

```

<210> 81  
<211> 1172  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 6803876CB1

<400> 81  
gaagagccgg gtggcagctg acagaggaag ccgctccaat accttcacca taaatagtgg 60  
ccatatatat atagtttaag aaggctctcc atttggcatc gtttaattta tatgttatgt 120  
tctaagcaca gcttctcttc tcctattttc atcctgcaag caactccaaa tatttgaaat 180  
aaagttttaca ttgtagttat tttccaatct ttgcttgata agtattaaga aatattggac 240  
ttgctgccgt aatttaaagc tctgttgatt ttgtttccct ttggatTTTT gggggagggg 300  
agcactgtgt ttatgctgga atatgaagtc tgagaccttc cgggtgctggg aacgcccgt 360  
gagccgcccg acggacgctc gtcttcgccc gccatggccg agagcgactg ggacacgggtg 420  
acggtgctgc gcaagaaggg ccctacggcc gcccaggcca aatccaagca ggctatctta 480  
gcggcacaga gacgaggaga agatgtggag acttccaaga aatgggctgc tggccagaac 540  
aaacaacatt ctattacca gaacacggcc aagctggacc gggagacaga ggagctgcac 600  
catgacaggg tgaccctgga ggtgggcaag gtgatccagc aaggtcggca gagcaagggg 660  
cttacgcaga aggacctggc caccgaaatc aatgagaagc cacaggtgat cgcggtactat 720  
gagagcggac gggccatacc caataaccag gtgcttgcca aaatcgagcg ggccattgat 780  
gtgggaacca ggagtgtctg tgtgtgaga gccagtgag gcacgggagc ggcgtggccc 840  
cccttccctg ggaagatgtt gaagaggggc agcgttgggg ggaggagggg gctggtgctg 900  
cagtggccag tggctgtggc tcatttcagg cctcaagctc cggggaaagg acattggaaa 960  
gcccatcgag aaggggccta gggcgaaatg aacacaaagc ctcgaaatca gtgcgctcca 1020  
gctgatctcg ttccgcccgt tccccttggc cgccagtcc gttctcctca cgggccgaac 1080  
ggaacaaggg gtccagcttg cgggggaccc tccccagccc attcctgctg tcaaacaac 1140  
aaaacctgac aaagcgcaaa aaaaaaaaaa aa 1172

<210> 82  
<211> 3424  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 7506281CB1

<400> 82  
agtgactgcg cggcccggacc cgggtcgggtg actgcggcaa gaccctggg tccccgcgcg 60  
gacgcatccc agtctgggcg ggacgctcgg ccgcgccgag gcgggcaagc ctggcagggc 120  
agagggagcc cgggctccga ggttgccttt cgccccgag gatcagtctt ggccccaaag 180  
cgcgacgcac aaatccacat aacctgagga ccatggatgc tgatgagggc caagacatgt 240  
cccaagtctt agggaaaggaa agccccctg taagcgatac tccagatgag ggcgatgagc 300  
ccatgccgat ccccgaggac ctctccacca cctcgggagg acagcaaagc tccaagagtg 360  
acagagtcgt ggtcattaaa gaagaaacta atcacagtga aatggcagaa gacctgtgca 420  
agataggatc agagagatct ctctgtgctg acagactagc aagtaacgct gccaaacgta 480  
agagctctat gcctcagaaa tttcttgggg acaagggcct gtccgacacg ccctacgaca 540  
gcagcgccag ctacgagaag gagaacgaaa tgatgaagtc ccacgtgatg gaccaagcca 600  
tcaacaacgc catcaactac ctgggggccc agtccctgcg cccgctgggt cagacgcccc 660  
cggcggtgtc caggtgtggt ccggctcatc gcccgatgta ccagctgcac aagccgctcg 720  
cggaggggac cccgcgtctc aaccactcgg cccaggacag cgccgtggag aacctgctgc 780  
tgctctccaa ggccaagtgt gtgccctcgg agcgcgaggg gtccccgagc aacagctgcc 840  
aagactccac ggacaccgag agcaacaacg aggagcagcg cagcgggtctc atctacctga 900

```

ccaaccacat cgccccgcac gcgcgcaacg ggctgtcgct caaggaggag caccgcgcct 960
acgacctgct gcgcgcccgc tccgagaact cgcaggacgc gctccgcgtg gtcagcacca 1020
gcggggagca gatgaagggtg tacaagtgcg aacactgccg ggtgtctctc ctggatcacg 1080
tcatgtacac catccacatg ggctgccacg gcttccgtga tccttttgag tgcaacatgt 1140
gcggctacca cagccaggac cggtagcagt tctcgtcgca cataacgcga ggggagcacc 1200
gcttccacat gagctaaagc cctcccgcgc cccacccca gaccccgagc caccacagga 1260
aaagcacaag gactgccgccc ttctcgctcc cgccagcagc atagactgga ctggaccaga 1320
caatgtttgt tttggatttg taactgtttt ttgttttttg tttgagttgg ttgattgggg 1380
tttgattttgc ttttgaaaag atttttatatt ttagaggcag ggctgcattg ggagcatcca 1440
gaactgtctac ctctcctagat gtttccccag accgctggct gagattccct cacctgtcgc 1500
ttcctagaat ccccttctcc aaacgattag tctaaatttt cagagagaaa tagataaaac 1560
acgccacagc ctgggaagga gcgtgctcta ccctgtgcta agcacggggt tcgcgcacca 1620
ggtgtctttt tccagtcccc agaagcagag agcacagccc ctgctgtgtg ggtctgcagg 1680
tgagcagaca ggacaggtgt gccgccaccc aagtgccaaag acacagcagg gccacaacc 1740
tgtgcccagg ccagcttcca gctacatgca tctagggcgg agaggctgca cttgtgagag 1800
aaaatactat ttcaagtcac attctgcgta ggaaaatgaa ttgtttgggg aaagtctgt 1860
ctgtcagact gccctgggtg gagggagacg ccgggctaga gcctttggga tcgtcctgga 1920
ttcactggct ttgcggaggc tgctcagatg gcctgagcct cccgaggctt gctgccccgt 1980
aggaggagac tgtcttcccg tgggcatatc tggggagccc tgttccccgc ttttccact 2040
ccataccttt aatggccccc aaaatctgtc actacaattt aaacaccagt cccgaaattt 2100
ggatcttctt tctttttgaa tctctcaaac ggcaacattc ctcagaaaac aaagctttat 2160
ttcaaatctc ttcttccct ggctggttcc atctagtacc agaggcctct tttcctgaag 2220
aaatccaatc ctagccctca ttttaattat gtacatctgt ttgtagccac aagcctgaat 2280
ttctcagtgt tggtaagttt ctttacctac cctcactata tattattctc gttttaaaac 2340
ccataaagga gtgatttaga acagtcatta attttcaact caatgaaata tgtgaagccc 2400
agcatctctg ttgctaacac acagagctca cctgtttgaa accaagcttt caaacatgtt 2460
gaagctcttt actgtaaaag caagccagca tgttgttcca cacatacata ggatggctgg 2520
ctctgcacct gtaggatatt ggaatgcaca gggcaattga gggactgagc cagaccttcg 2580
gagagtaatg ccaccagatc cctaggaaa gagggaggcaa atggcactgc aggtgagaac 2640
cccgcccatc cgtgctatga catggaggca ctgaagccc aggaagggtg gtggagattc 2700
taatcccaac aagcaagggt ctccttcaag attaatgcta tcaatcatta aggtcattac 2760
tctcaaccac ctaggcaatg aagaatatac catttcaa atttacagta cttgtcttca 2820
ccaacactgt cccaagggtg aatgaagcaa cagagaggaa attgtacata agtacctcag 2880
catttaatcc aaacaggggt tcttagtctc agcactatga cattttgggc tgactactta 2940
tttgtaggc gggagctctc ctgtgcattg taggataatt agcagtatcc ctggtggcta 3000
cccaatagac gccagtagca cccgaattg acaacccaaa ctctccagac atcaccaact 3060
gtcccctgcg aggagaaatc actcctgggg gagaaccact gacccaaatg aattctaaac 3120
caatcaaagt tctgggaagc cctccaagaa aaaaaataga aaagcacttg aagaatattc 3180
ccaatattcc cggtcagcag tatcaaggct gacttgtgtt catgtggagt cattataaat 3240
tctataaatc aattattccc cttcggtctt aaaaatatat ttctcataa acatttgagt 3300
ttgttgaaa agatggagtt tacaagata ccatcttga gtcattggatt tctctgtca 3360
cagaagggtg tggcatttgg aaacgggaat aaacaaaatt gctgcaccaa tgactgagt 3420
gaag 3424

```

&lt;210&gt; 83

&lt;211&gt; 1882

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7506175CB1

&lt;400&gt; 83

```

ctttccagca gagcggggcg gggcggggct gagegtgttt acatccgccg ggtgcgcgcc 60
ttcgcgcgcc gaggtcgttc ggctcgggta ccatacctccg cgccatggac accagcgacc 120

```

tgttcgccag ctgcaggaag ggggatgtgg gccgagtgcg gtacctgctg gagcagcgag 180  
acgtggagggt gaattgtgcg gacaagtggg acagcaccac cttgtactat gcctgcttgt 240  
gtggggcacga ggagctggta ctctaccttc tggccaatgg agcccgtgc gaggccaaca 300  
ccttcgatgg tgagcgctgc ctctatgggg cactgagtga ccccatccgc cgggctctac 360  
gcgattacaa gcaggtcacg gcttcctgca ggaggcggga ttactatgac gacttcttgc 420  
agcggcttct agagcagggc atccacagtg acgtggctct ttagtagacac ggaagccat 480  
tccgggtgca tcgctgcgtc ctgggtgcac gtagtgccta ctttgccaac atgctggaca 540  
ccaaatggaa gggcaagagt gtcgtggttc tcaggcacc actgatcaac cccgtggcct 600  
ttggggccct gctgcagtac ctgtacacag tggcgtctaa gccaggcacg tgtgtgaagg 660  
tgtgacccat cgagccccc cctgcagacc cccgcctccg ggaggacatg gcgctgctgg 720  
ccgattgtgc cctgcccccc gagctccgag gtgatcttgg ggagctgccc ttcccttgc 780  
ctgacggctt caacagctgc cctgacatct gcttccgagt ggctggctgc agcttctct 840  
gccacaaggc ctttttctgt ggccgagtg actacttccg agccctgctg gatgaccact 900  
tccgagagag cgaggagcca gcgacctcag ggggcccccc agccgtcacc ctgcatggca 960  
tctcacccca cgtcttccact cacgtgctct actacatgta cagcgaccac actgagctgt 1020  
cccccgaggc agcctatgat gtgctgagcg tcgccgacat gtacctgctg ccaggcctga 1080  
agaggctgtg cgcccgacg ctggctcaga tgctagacga ggacactgtg gtgggtgtgt 1140  
ggcgcgtggc caagctcttc cgcctggcgc ggcttgagga ccagtgcact gactacatgg 1200  
ccaaagtcct tgagaagctg gtggagcggg aggacttctg ggaggcggg aaggaggagg 1260  
cagcggctgt ggcagcccg caggagacgg actctatccc gctgggtggac gacatccgct 1320  
tccacgtggc cagcacggtg cagacctaca gcgccataga ggaggcgcag cagcgtctgc 1380  
gggactcga ggacctgctc gtgtccatcg gtctggactg ttgagccct ggctgggcag 1440  
ccccaggggc caggagctct cttggagaca agcatgtgta tgcgtttgtg tgcagctctt 1500  
cttctgctc cctgcacatt gagggcttca tggggggtgc gaggggctca gtggggcttc 1560  
tcttccctcc atgagcctgg agacccagg ggaggatcca tttgggatga gccccctccc 1620  
cccaatgcac aagccagccc ccaagaccct ggggggtggac accactcagg gaaacctggg 1680  
gtgggggtgg gctttgtct tagcactttc cttctccaga tccccctac ccacccag 1740  
cccaaatcca gtctctggc ccttgccatg ccctgaattg cttctctaag ctgggtgttc 1800  
catgcacagg gccattcagg aagggtggg ggagtgtgtg tggcaataaa gcttgaaggc 1860  
accgtgggag caaaaaaaaa aa 1882

<210> 84

<211> 2223

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 7506303CB1

<400> 84

gtgcagcctg atgcgcaacg tggggactca ggcgcgctgg gcggcaggag ttgcttccgg 60  
ccgtgttgggt ggtctgaatt gagaagccgc gactaaggga agatggagac aatactggag 120  
cagcagcggc gctatcatga ggagaaggaa cggctcatgg acgtcatggc taaagagatg 180  
ctcaccaaga agtccacgct ccgggaccag atcaattctg atcaccgcac tcgggccatg 240  
caagatatct gtgtgccaat gtcagtggaa tttgaggaa tcctgaaggc tcgagagaat 300  
ccaagtgaag aggcacaaaa cttggtggag ttcacagatg aagagggata tggtcgttat 360  
ctcgatctcc atgactgtta cctcaagtac attaacctga aggcactctga gaagctggat 420  
tatatcacat acctgtccat ctttgaccaa ttatttgaca ttcctaaaga aaggaagaat 480  
gcagagtata agagatacct agagatgctg cttgagtacc ttcaggatta cacagataga 540  
gtgaagcctc tccaagatca gaatgaactt tttgggaaga ttcaggctga gtttgagaag 600  
aaatgggaga atgggacctt tcctggatgg ccgaaagaga caagcagtg cctgacccat 660  
gctggagccc atcttgacct ctctgcattc tcctcctggg aggagtgggc tctctgggt 720  
ttggacagat tgaatctgc tctcttagct ttaggcttga aatgtggcgg gaccctagaa 780  
gagcgagccc agagactatt cagtaccaa ggaaagtccc tggagtcaat tgatacctct 840  
ttgtttgcca aaaatcccaa gtcaaagggc accaagcgag acactgaaag gaacaaagac 900

```

attgcttttc tagaagccca gatctatgaa tatgtagaga ttctcgggga acagcgacat 960
ctcactcatg aaaatgtaca gcgcaagcaa gccaggacag gagaagagcg agaagaagag 1020
gaagaagagc agatcagtga gagttagagt gaagatgaag agaacgagat catttacaac 1080
cccaaaaacc tgccacttgg ctgggatggc aaacctattc cctactggct gtataagctt 1140
catggcctaa atatcaacta caactgtgag atttgtggaa actacaccta ccgagggccc 1200
aaagccttcc agcgacactt tgctgaatgg cgctcatgctc atggcatgag gtgtttgggc 1260
atcccaaaca ctgctcactt tgctaattgg acacagattg aagatgctgt ctcttgtgg 1320
gccaaactga aattgcagaa ggcttcagaa cgatggcagc ctgacactga ggaagaatat 1380
gaagactcaa gtgggaatgt tgtgaataag aagacatacg aggatctgaa aagacaagga 1440
ctgctctagt gttgagggat gtagctcagc ttttgggcta gccaggctt ccctaagatc 1500
tgctttttct atttctccca accaaatcct cttaaagacc ctttgctatg tagtctcatg 1560
gtctagcatg catctttag tagaaaggca tgctggcaga ttgcagggtt gagatgtgtt 1620
ttatctgttt tatattttaa aagattctgc cagaaaataa aaccagacct tgttctaaag 1680
cccagggtta tggaccaact cagtgttca ggtcttaatg cctccatacc tcttcctcac 1740
caactttact agtagctgag atttaatggg cacctattat gctacataac atgttaggta 1800
aatctgacct gacctcttcc cccaccctcc tttgttgctg cttccctgaa tgagtattac 1860
ccaggatga ggtctgccat cagcttagtt agccattgat gcaaatacta gggaaagact 1920
aggaggatga gccagggttg ctactaagga ctaagtgtcg caccaagggt tgcttttgt 1980
atttgcataa agaaaggagt tggagctggg tgcagtggct tgtgcctgta gtcccagcta 2040
cttgggagggc tgaggcagga gggttgcttg agactagcct aggtaacata gtgagacct 2100
gtctcattaa aaaaaaagg catggtggca cgcactgtag tcccagctac tcaggagact 2160
gaggctagaa gatcctttga acctaggagt ttgagaccag cctgggagat atagtaggc 2220
ccc 2223

```

&lt;210&gt; 85

&lt;211&gt; 956

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7353336CB1

&lt;400&gt; 85

```

agcggaatga acggcggcgg gaggtgaaat cgggttctaa cgggtccggg gctcccagcg 60
ctataaaaac ttataaaacc ccccgagacc cgagcagtggt gaagaagagg cgagaacgac 120
ccccggaccg accaaagccc gcgcgccgct gcatcccgcg tccagcacct acgtcccgc 180
gccgtcgccg ccgccaccat gcccaagaga aaggctgaag gggatgctaa gggagataaa 240
gcaaagggtga aggacgaacc acagagaaga tccgcgaggt tgtctgctaa acctgtcct 300
ccaaagccag agcccaagcc taaaaaggcc cctgcaaaga agggagagaa ggtacccaaa 360
gggaaaaagg gaaaagctga tgctggcaag gaggggaata accctgcaga aaatggagat 420
gccaaaacag accaggcaca gaaagctgaa ggtgctggag atgccaaaac agaccaggca 480
cagaaagctg aagggtgctgg agatgccaag tgaagtgtgt gcatttttga taactgtgta 540
cttctggtga ctgtacagtt tgaaatacta ttttttatca agttttataa aaatgcagaa 600
ttttgtttta cttttttttt ttttttaaaa gctatgtggt tagcacacag aacacttcat 660
ggtggttttg gggggaaggg gcatatgtca ctaatagaat gtctccaaag ctggattgat 720
gtggagaaaa cacctttccc ttctagtttg gagagacttc ctctgggctc ccaggaggag 780
ggattccctg actttgacac acatggccac ctgtggcaca aaagcctggg ggtatagaaa 840
aacaaattgg tgtttatgtc ctcttctccc tttcatcttt cagcatagac ttaatccctt 900
taagccccag actctggttg agacctgacc ctagtcatgg gttccccgtg tgtcgg 956

```

&lt;210&gt; 86

&lt;211&gt; 2633

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 3001652CB1

&lt;400&gt; 86

```
tgaggaaaga caggaaagcg ccagcatctc caccttcccc ggaagcctcc ctttgccagg 60
cagaaagggg tccccatggg gccgcccctg gcgcgcgcgc cggcccacgt acccggggag 120
gccgggcccc ggaggacgag ggaagcagg cggggcgccg tgagcttcgc ggacgtggcc 180
gtgtacttct ctcccagga gtgggaatgc ctgcgccag cgcagagggc cctgtaccgg 240
gacgtgatgc gggagacctt cggccacctg ggcgcgctgg gattttcagt tcccaaacc 300
gcttttatct cgtgggtgga aggagaagtg gaggcgtgga gcccgagggc ccaggatccc 360
gacggtgaga gctctgcagc ttccagcagg ggccaaggac aggaagcagg atccagggat 420
gggaatgagg agaaggaag gctgaagaag tgtccaaac aaaaagagggt ggcgcatgaa 480
gtggctgtca aggagtgggt gccagcgctc gcctgccag agttctgcaa ccctaggcag 540
agccccatga atccctggct caaggacact ctgaccgaa gactgcccc ctcttgcccc 600
gactgtggcc gcaacttcag ctacccttcc ctctggcca gccaccagcg ggtccactcc 660
ggggagcggc ccttctcctg cggccagtgt caggcgctgt tctcccagcg cagggtacct 720
ctccagcatc agttcatcca caccggcgag aagccctacc cctgccccga ctgcgggcgc 780
cgcttcgcgc agagggttc cctggctatc cacaggcggg ctacacccgg ggagaagcct 840
tacgctgtct cagactgcaa gactcgcttc acttaccct acctgctggc catccaccag 900
cgcaagcaca cgggcgagaa gccctacagc tgccccgatt gcagcctccg tttcgctac 960
acctccctgc tggccatcca caggcgcata cacaccggcg agaagcccta cccctgtcct 1020
gactgcggcc gccgcttcac ctattcttcc ctctcctca gtcaccggcg cattcactcc 1080
gacagccggc ccttcccctg cgtggagtgt gggaaaggct tcaagcgcaa gaccgcccgt 1140
gaagcccatc ggtggatcca ccgctcctgc agcgagaggc gcgcgtggca gcaggccgtg 1200
gtggggcggt cagagcccat ccctgttttg ggaggcaagg atccccagt tcaattccgg 1260
cactttccag atatatttca agagtctgt caacagaggc ttcaggaccg cggggtccct 1320
tcaaatgccc cgcgatccc agggcaatca ccgcgcagct tcttccggga tcgtcgccaa 1380
tcctggcgcc ttgcgtattg cgggtacaga ggggtaagt aagcttcggg cccttacatc 1440
tttcttgaag gcaagaagcc tctctctac ttccagaca ccccgccctc cccactagaa 1500
aaagcagccg aagcggtttt atttaagggc aagtgggacg atgaggccag agaaatggcg 1560
ccgccccag ccccgctcct ggcgcgagg cccggggaga cccggcctgg ttgcaggaag 1620
cccgggactg tgagcttcgc ggacgtggcc gtgtacttct ccccgaggga gtggggctgt 1680
ctgcgccccg cgcagagggc tctgtaccgg gacgtgatgc aggagacct cggccacctg 1740
ggcgcgctcg gattcccagg ccccaaacca gccctcatct cttggatgga acaggagagt 1800
gaggcttgga gccccgccgc ccaggatcct gagaaggggg aaagactggg aggaagctcg 1860
agaggagatg tcccaaacag gaaggaagag gaaccggagg aagtcccaag agccaaaggg 1920
cctagaaagg ctctgtgtaa ggagagtcct gaagtgtgg tggaaacgaa ccctgaccca 1980
gctattagcg tggccccggc acgggcacag ccacccaaaa atgctgcctg ggaccgcacc 2040
acaggagcac agccccggc accataccc agcatggatg ctacggccgg ccagcggcgc 2100
cacgtgtgca cggactgcgc ccgcgccttc acctaccct cactgctggt cagccacagg 2160
cgcatgcact cgggggagcg gcctttcccc tgccccgagt gtggcatgcg cttcaagagg 2220
aagttcgcag tggaaagcga ccagtggatc caccgctcct gctccggggg gcggcggggc 2280
cggaggcctg ggatccgggc tgtgcctcgg gccccgtcc gaggtgaccg ggaccgcct 2340
gtgctcttcc ggcactacct agacatcttc gaggagtgcg gctgagcggc accgcaggct 2400
ggagttgagc ctgaccttg caccgaaggac tgacggatcc ctgaggtggg ccactgagtc 2460
ggggactccg gaactgaaat tcatgccctg ggctttctc aaggatccct caagtttcca 2520
acttgtaaaa agaaaagtgc ctgtaagat tcgaatagat tagacttgcc acccatctcc 2580
ccagtctttt gtttaacaaa aaaaaaaaaa aagggccgct cgcatctag aac 2633
```

&lt;210&gt; 87

&lt;211&gt; 4521

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1689128CB1

&lt;400&gt; 87

```

gcatctagaa ccacacaccg gagaaggggt ttccccctgc ccgcccggcc gcccccaac 60
tacctccccg ctccctgccag tgtctgcctc tctgccccca cgggggttta ttgatctca 120
cattaaccaa ggaggagaat gtgagaaaag ggggaaaatg cccaggagga agcaggagca 180
gccaagcgc cttccctcac atgttagtag gcaggaagag gcgaggaggag agctcagtga 240
aggagaacac tggtaggaa actcttcaga gactccgtca gaagcatcct atggggaaagt 300
ccaggagaac tataagttgt ctctggagga ccggatccag gagcagtgca cgtccccgga 360
cacctccttg ggaagcacga ctcccagcag ccacacgttg gagctggtgg cactggacag 420
tgaggtcctg cgagactcac tgcagtgtca agatcacctt tccccgggag tgtcttcttt 480
gtgtgacgat gacccaggct ccaacaagcc cctgagcagc aatttgaggc ggctgctgga 540
ggctgggttc ctcaaactcg atgtgcagc cacggcta atggcagagtgg agtccccctg 600
aaacgttggc tcgaacctct ccttttcccc gccttcccac cagccccagc agctcagtgt 660
tctggccagg aagtggccg agaagcagga acagaatgac caatacactc caagtaacct 720
ttttatatgg aatcaaggta agtggttgcc aaactcaacc accactgca gcttctctcc 780
ggattcagcc atcctaaaac tgaagctgc agccaatgcc gttctgcagg acaaatctct 840
caccaggact gaggagacca tgcgatttga gtccttttcc tcccccttta gctcccagtc 900
tgccagttct accttggccg ctttgtccaa gaaggtcagt gaaagaagtt tgactcctgg 960
tcaggagcac cctccaccgg ccagctcctt cctttccctg gcttctatga cctcctcagc 1020
ggcccttctg aaggaggtgg ccgcaagggc tgcgggcagt cttctggctg agaaatcatc 1080
gctgctgcct gaggaccctc taccgcccc gccttcagag aagaaaccag aaaaagtcac 1140
tccgccacct ccaccgccac ctccaccacc tccaccacca ccaccacaat ccctggaatt 1200
attattactc ccagttccta agggaagagt ttctaaaccc tccaattcag cctcagaaga 1260
ggaaagtgga aagcctttcc agtgtccaat atgtggtttg gttattaaaa ggaagagtta 1320
ctggaagcgg cacatggtga ttcacacagg tttaaaaagt catcagtgtc cgctctgtcc 1380
attccggtgt gctcgcaagg acaatctcaa atcccacatg aaggttcac agcaccagga 1440
tcggggagag acctttcagt gccagctgtg cccttttact tctcacgc ccctcagcct 1500
gaaactccac atgcgttgcc atcagcactt cctgaggaca gaagccaagg tgaaggagga 1560
gatcccagac ccagatgtca agggatctcc ccacctcagt gacagtgcct gcctggggca 1620
gcaaagggaa ggaggaggga cagagctagt ggggaccatg atgacgtcta aactccaga 1680
gaggactagc caggaggagg ctggcgtctc gcctttgctg gtgaaggagg aaccaagga 1740
agataacggc ctgcccacct cctttacttt gaatgctgcc gacaggcccc ccaaccacac 1800
aaagctgaaa gacccctccg agtatgtggc caacagtgcg tcagcattgt tcagccagga 1860
catctctgtt aagatggcgt ctgattttct catgaagtgc tcagctgcaa atcagaagga 1920
gcccatgaat cttaatttta aagtgaagaa ggagcctaag gaaggggagt ccctaagcac 1980
gactttgcct cgggtccagct atgtgttcag cccggaatct gaagtgtcag cccagggcgt 2040
ctctgaggac gcactaaagc cccaggaagg gaagggaagt gtgctaaggc gggatgtgtc 2100
agtcaaagca gcctctgaac ttctcatgaa actctcagcg gaaagctaca aggaaacaca 2160
gatggtgaag attaaagagg aacccatgga gggtgacatc caggactccc atgtctcgat 2220
atcaccacgc cggaatgttg gctacagcac tttaatcggg cgagagaaaa ccgaaccctt 2280
acagaagatg ccagagggca gactaccccc agagagaaac ctcttcagtc aggatatctc 2340
tgtgaaaatg gcttccgagc tctcttttca actgtcagaa aaagttagca aagagcacia 2400
tcatacaaaa gaaaacacca tccggaccac gaccagccct ttcttttcag aagacacatt 2460
tagacaatca ccattcacct ccaattcaaa agaactgtgc cccagtgcct ccgtgctgca 2520
cggaagaata tcagctccag aaacagaaaa gatagtccta gaggcaggaa atggattacc 2580
atcctggaaa ttcaatgacc agcttttttc ctgtgacgtg tgtgggaaag tgtttggccg 2640
acagcagaca ttgtcccgac acctctcgct gcacacagag gaaagaaaat acaaatgcca 2700
cttgtgcccc tatgtgcta agtgccgtgc aaatctgaac cagcacttga ccgtccattc 2760
cttggaagctg gtgagtacag acaccgagga cattgtcagc gccgtcacct ctgaaggcag 2820
tgatgggaag aaacatcctt attattacag ttgtcacgtg tgtggatttg agaccgagct 2880
caatgtccag tttgtcagcc acatgtcact ccacgtggac aaggagcagt ggatgttttc 2940
gatctgctgc actgcctgcg acttcgtcac catggaggaa gcagagataa agactcacat 3000
tggcaccaag cacacagggg aagacaggaa gacccccagc gaatcaaata gccctcttc 3060
atcctccctc tcagctctga gtgattcagc caacagcaaa gatgattcag atggctccca 3120

```



```

gaaaaacaag ggcgggaaca atctgctggt catctctgtc atgcctggga gccagccctc 3180
actgaacagt gaggaaaagc cagagaaagg gttcgaatgt gttttttgca actttgtctg 3240
caagacgaag aacatgtttg agcgtcatct gcagatacac ctcatcaccg ggatgtttga 3300
gtgtgatgtg tgccacaagt tcatgaagac ccccgaaacg ctgctggagc ataagaaatg 3360
ccacactgtc cccaccgggt ggctcaactc aggacagtgg tgagtttcag actcctctag 3420
gtgcccattc tgcattttatt ccaccaaccg ccccgctgcc atggagtgcc acctcaagac 3480
ccactacaag atggagtaca agtgccggat ctgccagacg gtgaaggcca accagctgga 3540
gctggagacg cacaccggg agcaccgcct gggcaaccac tacaagtgcg accagtgcg 3600
ctacctgtcc aagaccgcca acaagctcat cgagcacgtg cggtccaca ccggggagcg 3660
gcccttccac tgtgaccagt gcagctacag ctgcaagcgc aaggacaatc tcaacctgca 3720
caagaagctg aagcacgccc cacgccagac cttcagctgc gaagagtgcc tgttcaagac 3780
cacacaccct ttcgtcttca gccgccacgt caagaagcac cagagtgggg actgccccga 3840
ggaggacaag aagggcctgt gtccagcccc caaggaaccg gccggcccg ggccccgct 3900
cctggtggtc gggagctccc ggaatctcct gtctcccctg tcagtattgt ctgcctccca 3960
ggctctgcag accgtggccc tgtcggcagc ccacggcagc agctcagagc ccaacctggc 4020
actcaaggct ttggccttca acggctcccc tttgcgcttt gacaagtacc ggaactcaga 4080
ttttgcccac ctcatctcct tgacaatgtt ataccceaag aaccacttgg atctcacatt 4140
ccaccctccc cgacctcaga ctgcgcctcc cagcatcccc tcacccaaac actccttccc 4200
ggcctatctc ggactgagag aaagagcaga gactgtctga gggcagccat gttctgtacc 4260
aaaaacagag agacaaaaga caaaaaaaaa aaaaaaaacc aaaaaactta aacacaaccc 4320
cagcaggtgt atgttgctgc aaaacctaca gaccccgatg ggtctgtgaa catgtgtact 4380
gtaactctcc tagtaaggaa tagaaaattg gctctgtgtg tatacctatt gcattgacct 4440
gaaagcggct ttatccaatc ttcagagagg tgacctactg gatacttcta ccttcagagg 4500
catgctcccc caggcaccaa g                                     4521

```

&lt;210&gt; 88

&lt;211&gt; 3907

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2362969CB1

&lt;400&gt; 88

```

ttttgcggcc cccctttttg tgtatttttt gtctttgagt gacaccagag ctgggttcac 60
ggtgatggct ctcccataat tgtgggcttt gcttgaatgg gttttaacgg aggactcttt 120
ggttttagcct gaacaacagt tgttgtagag gaggccagtc caggggtattg tcccagtggt 180
ggtcgttgta aagggttgag tgggtatagc gattagctgc gggggagagg ggcacggggc 240
agaggggaaga gagcaaggtc tgctgtgct gccccgggct ccgaggccag cttcacagag 300
tccagagggc tgcagaataa gaacagaggg ggggccaatg ggaaaggagg gcggggcagc 360
ctcaatgcc a gcgacgaag gacaccccca aattgtgtg ctgaggatat caaagccagc 420
ccttctccca ccaacaaaag gaaaaacaag cctccaatgg agctggacct gaactccagc 480
tctgaggaca ataagcctgg aaagcgtgtc cgcacaaatt ccagaagcac tcccactacc 540
cctcaaggga aaccagagac tacttttttg gaccaaggct gctcttctcc agtggttaac 600
gactgtcccc acccaaactg caacaaaaag tacaagcaca ttaacggcct gaggtaccac 660
caggctcatg cacacttaga cccagaaaac aagctggagt tccagcctga cagtgaggac 720
aagatctcgg actgtgagga aggattgagt aatgtggcac ttgaatgcag tgagccaagc 780
acaagtgtat ctgcttatga ccagttgaag gcaccggcat cccctggtgc tggaaaccca 840
cctgggaccc caaagggaag gagagagctg atgagcaatg gccaggttc cattattggt 900
gctaaagctg ggaagaattc tggcaaaaag aagggcctta acaatgaact gaacaacctt 960
ccagtaatct ccaacatgac ggctgcgtta gacagttgct cggcagcaga cggcagttt 1020
gctgctgaga tgcctaaact ggaagcagaa ggattaattg acaagaaaaa ttaggagat 1080
aaagaaaagg gcaaaaaagc taccaactgc aaaacggaca aaaacctctc taaactgaaa 1140
agtgcggcgc ccattgcccc tgccccagcc cccactcccc cgcagctaat cgctataccc 1200
actgcaacct ttacaacgac caccactggg acaatacccg gactgccctc cctcacaaca 1260

```

```

actgttggtc aggctacacc aaagagtcct ccgttaaaac ccattcaacc aaagcccaca 1320
attatgggag agcccatcac cgtgaaccca gctctggtgt cactcaaaga caaaaagaaa 1380
aaggagaagc gaaagctaaa ggacaaagaa gggaaagaga cgggaagccc aaaaatggat 1440
gccaaagctg ggaaactaga ggactccaag ggggccagca aagatttacc tgggcatttt 1500
ttaaaggatc atctcaacaa gaatgaaggg ctggcaaatg gactgtcgga gtctcaggag 1560
agccgcatgg ccagtatcaa agctgaggcc gataagggtt acactttcac agacaacgct 1620
cccagccctt ccatagggag cgctcgagg ctggaatgca gcactttggt gaacgggcag 1680
gcaccaatgg cacctctgca tgtgttgacc cagaatgggg cagagagttc agctgcaaag 1740
acaagcagcc cggcctattc agacatatct gatgctgctg acgatgggtg ttctgacagc 1800
aggtcggagg gtatgagatc aaaggccagt tccccatcag atattatttc tagtaaggac 1860
agtgttgtaa aagggcattc ttcaactaca gcacaatcat ctcaactgaa agagtcccat 1920
tctccctatt accacagcta tgatccttat tattctccaa gttacatgca ccctgggcag 1980
gtcggtgccc ctgcagctgg aaatagtggg agcacgcagg gaatgaagat caagaaggag 2040
tctgaggaag atgctgaaaa gaaagacaag gcagagcagt tagattctaa gaaagtggac 2100
cacaattctg catccttaca gcctcagcac cagtcggtga tcacacaaag acatcctgcc 2160
ctggctcagt cactttatta tggccagtat gcatatgggc tctatatgga ccagaagtct 2220
ctgatggcca ccagccctgc ctatagacag cagtatgaga agtactatga ggaccagagg 2280
ctggcagagc agaaaatggc ccagactggg agaggagact gtgaaaggaa aagtgcagtc 2340
cccttgaaag agctgggcaa ggaggaaact aaacagaaaa atatgccatc ggccacaatc 2400
tcaaaagctc cctctactcc ggagcctaac aaaaaccatt ctaaactagg gccatcagtg 2460
cctaataaaa ctgaggagac aggtaaatcg cagcttctct ccaatcacca gcagcagctt 2520
caggccgaca gcttcaaagc taagcagatg gaaaaccacc agcttattaa ggaggctgta 2580
gaaatgaagt ctgtcatgga ctctatgaag cagacaggtg tagaccaac ctgcagattt 2640
aaacaagacc cagattcaag aacatggcat cattatgtat accagcccaa atatctggat 2700
cagcaaaagt cagaagaact tgatagagag aagaaattaa aagaggatag tccgagggaa 2760
actcctaata aagagagtgg tgtgccagc cttcctgtat cgtaacaag cattaaagag 2820
gagcccaaag aggccaaagc tcctgattct caatcaatgg aggagagcaa gctgaaaaat 2880
gatgatcgaa agactcctgt gaactggaag gactctcggg gaacaagagt ggctgtctcc 2940
tcacccatga gtcagcatca gtcatacata cagtacttgc atgcttatcc ttaccacag 3000
atgtacgacc ccagccatcc tgcataccgg gctgtttctc ccgtcctaata gcacagttat 3060
cctggggcat atctctctcc aggttttcat tatcctgttt atgggaagat gtcaggggaga 3120
gaagagacag agaaagtcaa taccagccct agcgtcaaca cgaaaacaac cactgaatct 3180
aaagcactgg atttgctcca gcagcatgct aaccaatacc gcagcaagtc tcctgctcct 3240
gtggaaaagg ctacagctga gcgggaacgg gaggcagaaa gggaaaggga tcgccactcc 3300
cccttcggcc agcggcacct gcacacgcac caccacacc acgttggcat gggttaccgc 3360
ctaatacccg gtcaatatga cccttttcaa ggcttgacct ctgctgccct tgttgccctc 3420
cagcaggtgg ctgcccaggc atctgcatct ggaatgtttc ctggacaaag aagagaataa 3480
caagattgga catgtacatt gtcattgtgac tggatcacat ggggaagcgc ttatatacaga 3540
catcagttcc atggtgttaa ttgaaatgc cttaatggca ggcaaaaacc agtcatttca 3600
tttttgtaaa ttacagattt tatcacagag taacaaatgt tgctgtaatt aaacttctgt 3660
tttatatata tatacgtata catatctgta tatatatata taaatataaa 3720
tatattcttt actttttgta tcagtaacca ggctcgaca cacagggtct gctgccacgt 3780
cgcaaaaccac tcagtaaaag aaattaaaaa tgtatttgtc atgttaaaaa gtgttagcca 3840
caaaaggcta cttactttct ttctcttttc ctttaaaatt gtaaatagat aatgttatgt 3900
atcaggg 3907

```

&lt;210&gt; 89

&lt;211&gt; 1802

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 4753527CB1

&lt;400&gt; 89

```

caaggtcgag agagggaggc ggcggtgggg aggaggagga ggaggaggag caggcgcccc 60
ccatggccgc cgctatcacc cgacatggcc gacctggagg agctctcccc cctgagccct 120
ctgccccccg gcagcccggg ttccggcgcc ggggcccggg tgagccccc gagggaggag 180
aggaagagga ggaggaggaa gaggaggcgg aggccgaggc ggtggcgggc ctgctgctga 240
acggcgggcag cgggtggggg ggcggaggcg gcggcgagg agtggggggc ggcgaggcag 300
agacgatgtc ggagccgagc cccgagagcg ccagccaggc cggggaggac gaagacgagg 360
aggaggacga cgaggaggag gaagatgaga gcagcagcag cggcgggggg gagggaggag 420
gtagcgccga gagcctggtg ggcagcagcg gcgggagcag cagcgacgag acccgctcgt 480
tgagcccccg gcgcgccagc agcagcagcg gggatgggga cggcaaggag ggcctggagg 540
agcccaaggg accgcggggc agccaggcg gcggcggggg cggcagcagt agcagcagcg 600
tagtctccag cggcgggcag gagggtacg ggactggggg aggcggaagc agcgcgacct 660
ccggggggcgg gcggggcagc ttggagatgt cgtcgatgg ggaacccctg agccgcatgg 720
actcgaggga cagcataagc agtactataa tggatgtaga cagcacaatt tccagtgggc 780
gttcaactcc agcaatgatg aatggacaag gaagcactac ttcttcaagc aaaaatattg 840
cctataattg ttgttgggac cagtgccagg cttgcttcaa ctctagccca gatctggcag 900
atcacatccg ttccatacat gtagatggtc agcaggagg ggtatttgtt tgcttatgga 960
aaggttgtaa agtatataac actccatcta ccagtcaaag ttggttacia aggcataatgc 1020
tgacacacga tggagacaaa cctttcaagt gtgttgttgg tggctgcaat gccagctttg 1080
cttctcaggg agggctagct cgtcatgtac ccacacactt cagtcagcag aactcctcaa 1140
aagtttctag ccagccaaag gccaaagaag aatctccttc taaagctgga atgaacaaaa 1200
ggaggaaatt aaagaacaaa agacgacgct cattaccacg gccacatgat ttcttcgatg 1260
cacaacact ggatgcgata agacatcgag ccatatgctt taacctctca gctcatatag 1320
aaagtttagg gaaggacac agtgttgttt ttcatagtac tgtaatagct aagagaaaag 1380
aagattctgg gaagatcaaa cttttgcttc attggatgcc tgaagacatt ctgcctgatg 1440
tgtgggtgaa tgaaagtgaa cgacatcagt taaaaactaa agtagttcat ttatcaaagc 1500
tacccaaaga tactgccttg cttttggacc caaacatata cagaacaatg ccgcagaaga 1560
ggttgaagag atttgacatc ctcaatttcc caaggtgatg agtgcttctt catctggaac 1620
ctcacagcca tgagaacgga aaggggacga cttattttta cgggttacag cgtaaacaca 1680
catctccaca ggtgtctcct ggctgctatt gcctaaggt cttgggttac ctcacctggt 1740
ccatcgtgtg gtttccctag aaccagaaa tttatctaag tgagactctc aggatagact 1800
ta

```

&lt;210&gt; 90

&lt;211&gt; 2469

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 6928688CB1

&lt;400&gt; 90

```

cggctcgagg cagtgggagc aggtcatcgg tctgttgagc atttttgtca tctgcaaata 60
attttagcca ggtggcaagt gtatggactg tggagtaaga agtacctgcc tctgcctctg 120
acttgggttc tggcctatag gaacaagaca ttttcagtaa tggaagaagg agcttccagg 180
ctggggaaga tcatcatttc cagtggtagc ttgacctgag atgataacctg accaccactt 240
ctgggaaggc ccagaataga cagaagagag agaacatact agagttgcta gcccctaaga 300
gttgaaacgaa atgggcaagg ccagctgaa ataacaggaa acacatgaaa aatgccacga 360
ttgtaatgtc tgtgaggaga gagcaaggca gttcttcagg ggaggggtca ttgtcattcg 420
aagatgtggc tgtgggcttc accaggagg agtggcagtt tttggaccag tctcagaagg 480
tctttgacaa ggaagtaatg ttggagaact acatcaacct agtatcaata gggatcagag 540
gcaccaagcc agattcgctc ttcaagtgg agcaaggaga acccccagg atagcagaag 600
gagcagccca cagtcaaadc tgtccagggt ttgtaatcca gagtagaaga tatgcaggaa 660
aagattctga tgcatttggg ggatatggga gatcatgcct ccatatcaag cgtgacaaaa 720
ctcttactgg agttaaatc catagatgtg ttaaaccag cagccctaaa tcacagctca 780
atgacctaca aaaaatttgt gcaggaggga aaccacatga atgcagtgtg tgcgggagag 840

```

```

ccttctccag gaaagcacag cttattcaac atcagagaac tgaaagagga gagaaacccc 900
atggatgtgg tgaatgtggg aaaacattca tgaggaagat tcagctcact gagcatcaga 960
gaactcacac aggagagaag ccccatgaat gtagtgaatg tggaaaagcc ttctccagaa 1020
agtcacagct catggtccat cagagaaccc atacaggaga gaaaccctac agatgcagcg 1080
aatgcggaag agccttcagc cggaaagtgc ggctcaatag acatcagcga tcacatactg 1140
gagagaaact ctatgggtgc agtgtgtgtg ggaaagcctt ttctcagaag gcataacctca 1200
ctgcacacca gagacttcac acaggagata agccttataa atgcagtgat tgtggaagaa 1260
ccttctatct taagtcagac ctgaccagac atcagaggat tcatacggga gagaaacctt 1320
atgaatgtag tgagtgtgaa aaagccttta gaagcaagtc gaagctcatt cagcatcagc 1380
gtactcacac tggagagaga ccatattcat gcagagaatg tggcaaagcc tttgccaca 1440
tgtcagtcct cattaaacat gagaaaactc acataagaga gacagccata aattcactga 1500
cggtaggaga accttcctca aggagtcata cctccttata catgagcgaa ctcatacaag 1560
agcaaaagac tgtgaacaca gtacctatag aaatgccttc ctcaggaacc ccgccattgt 1620
taaacaagag tgagcgcta gtgggcagaa atgtagtgat tgtggaacaa ccttttccaa 1680
gaaatcaagc ctttgtagtt aatcaggaat ttgaacagag aataagcctc acaaatgaag 1740
tgaatgtggc cccatcagta ataaattata tcttgtatct tacagatatt gtatcagaat 1800
aaatcttctg ataccagaaa ggtgggaaaa tcttttagtag gaatcctcca gccattata 1860
cccagagagc tcatttaggg ctgaaacact gtggaggcag tggttgtgga ggacatttct 1920
gtgagaagtc aaacctgtta aacaccattg aaatcatgtt gggtagaaat gcagctgtta 1980
tgaggaagtc tttacctgga gatgatctct caattggtat caaaagaatt cttacaggaa 2040
tatttatgaa agattgtgat ggtgctttta gcgctaaatt gtgcctcgtg tgccaaagaa 2100
acatagacaa ctgtagaggc aatggatatg gaaaacattt ttcgtaaaat atgggagaa 2160
ttatagaaaa gaaaaactct tgaaagtaag aaatgatggc aactcaccaa attaaactcat 2220
gtatcatact tttatgtctt gagagtcttg gagggacgtg tttaatgaaa tcttgaaccc 2280
caggaatagt attatgtaat ggaaacctat gaatatacaa atatggagta aaatgtgtta 2340
gttcttatga gtgtgaccaa attacacaaa tttagggtta aaaacaaatc tttcttcagt 2400
cggggggggg ccacaggttc aggggaacaa aggggaaggg gggatggggg gggagggttg 2460
ggcacatag

```

&lt;210&gt; 91

&lt;211&gt; 3878

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7506388CB1

&lt;400&gt; 91

```

tcccccttc ccctccctcg ctgcccctgg cccagcggga gcccccccc cccccccagt 60
gcgcctgtgc ggaggccctc ttgattatgt gtgccccttc cgggcgcccg cgttaccccc 120
gggtggagggt ggggaggga gacgctgagg agggaggaga ggcgaggag gcggtggagg 180
ggaggtgggg ggaatcagca aggacatggc tcctgactcc tgtgcggaac gtgagtgact 240
gagcggcaaa gcccgatgg tctctagcga cgggcccgtg tactggagg acgaggtctc 300
ccatagtatg aaggagatga ttggaggctg ttgcgtttgc tcagacgaga gaggtgggc 360
cgagaacccg ctggtttatt gcgacgggca cggctgcagc gtcgcggtgc atcaagcttg 420
ctatggcatt gttcaagtac ccactggacc gtggttttgc aggaatgtg aatctcagga 480
gagagcagcc agagttagat gtgaactttg tccccataag gatggagctt taaaaagaac 540
agataatggg ggttgggccc atgtggtttg tgccctgtat attccagagg tacaatttgc 600
caatgtttcc acaatggaac caattgtttt acagtctgtt ccgcatgatc gttataataa 660
gacttgctac atttgtgatg aacaagggaag agaaagcaaa gcagccactg gtgcttgcat 720
gacatgtaat aaacatggat gtcgacaggc tttccatgta acatgcgctc agtttgccgg 780
actgctttgt gaagaagaag gtaatggtgc cgataatgtc caatactgtg gctactgtaa 840
ataccatttt agtaagctga aaaagagcaa acggggatct aataggtcat atgatcaaag 900
tttaagtgat tcttcctctc actctcagga taaacatcat gagaaagaga aaaaaaata 960
taaagagaag gacaaacaca aacagaaaca caagaagcag ccagaacctat cacctgcatt 1020

```

```

ggttccatcc ttgactgtta ctacagaaaa aacttatata agcactagca acaactctat 1080
atctggatca ttgaagcgct tggaagatac tactgcacga tttacaaatg caaatttcca 1140
ggaagtctct gcacacacct ctagtggaaa agatgtttca gagactagag ggtcagaggg 1200
caaagggaag aaatcttcag ctacagctc aggtcaaagg ggaagaaagc ctggtggtgg 1260
aagaaatcca ggaacaactg tgcagcagc tagccctttt cctcaaggca gtttttcagg 1320
aactccaggc agtgtaaagt catcttctgg aagttcagtg cagtctcccc aggatttcct 1380
gagctttaca gactcagatc tgcgtaatga cagttactct cactcccaac agtcatcagc 1440
aaccaaagat gtacataaag gagagtctgg aagccaggaa gggggggtaa atagttttag 1500
taccttaatt ggcctccctt caacctcagc tgttacttca cagcctaaaa gctttgaaaa 1560
ttcacctgga gatttgggta attccagcct tcctacagca ggatataagc gggctcaaac 1620
ttctggcata gaagaagaaa ctgtaaagga aaagaaaagg aaaggaaata aacaaagtaa 1680
gcatgggcct ggcagacca aaggaaacaa aaatcaagag aatgtttctc atctctcagt 1740
ttcttctgct tcaccaacat catctgtagc atcagctgca ggaagcataa caagctctag 1800
tctgcagaaa tctcctacat tgctcaggaa tggaagtta cagagcctca gtgttggtc 1860
atctccagtt ggttcagaaa tttccatgca gtatcgcat gatggagctt gccaacaac 1920
tacgttctca gagtgtctga atgcaatata caacggtatt tataacagca atgatgtagc 1980
agtatcggtt ccaaatgtag tatctggctc gggatctagt actcctgtct ccagctctca 2040
cttacctcag cagtcttctg ggcatttgca acaagtagga gcgctctctc cctcagctgt 2100
gtcatctgca gccctctctg ttgtacaac tcaggcaaact actctatctg gatcttctct 2160
cagtcaggca ccatctcata tgtatggcaa tagatcaaact tcatcaatgg cagctcttat 2220
agctcagttc gaaaacaatc aaacagatca agatcttgga gacaatagcc gcaacctagt 2280
tggcagagga agctcacccc gaggaagtct ctgcacagc tcccctgtaa gcagcttaca 2340
gattcgctat gatcaaccag gcaacagcag tttggaaaat ctgcctccag tagcagccag 2400
catagaacag cttttggaga ggcagtggag tgaaggacag caatttttac tagaacaggg 2460
tactcctagt gacatttttag gaatgctgaa gtcattacac caacttcaag ttgaaaaccg 2520
aagattagag gaacaaatta aaaacttgac tgccaaaaag gaacggcttc agttattgaa 2580
tgcacagctt tcagtgcctt ttccaacaat aacagcaaact cctagtccgt ctcatcaaat 2640
acacacattt tcagcacaga ctgctctac tactgattcc ttgaacagca gtaagagccc 2700
tcataatgga aacagctttt tacctgataa ttctcttctt gtattaaatc aggacttaac 2760
ctccagtggg caaagtagca gcagctcatc agctctttct accccacctc ctgctgggca 2820
gagtccggct caacaaggct caggagttag tggagttcag cagggtcaatg gcgtgacagt 2880
gggggcacta gctagtggaa tgcagcctgt aacttcacc attcctgccg tgtctgcagt 2940
gggtggaata attggagctt tgccaggtaa ccaactggca attaatggca ttgtaggagc 3000
tttaaattgg gttatgcaga ctctgtcac aatgtcccag aaccctaccc ctctcaccca 3060
cacaaccgta ccacctaatg caacacatcc aatgccagct acactgacta acagtgcctc 3120
aggactagga ttactttctg accagcaacg acaaatactt attcatcaac agcagtttca 3180
gcagttgtta aattctcaac agctcacacc agtacacagg caccctccact tcacacagct 3240
accaccaacc catttctcac catccatgga gataatgcaa gtcagaaagt agcaagactt 3300
agtataaaaa ctgggcctgt agctcaagag aaaagttgac acctgagaaa catctagaaa 3360
ttgcctatcc tgctgttcta gcacttcac tggctgcctt tgcagtcctt ttactacagc 3420
tatgaagaaa cgcaacaaga aactcaatgc acaacaaagg attaatgtct gcaaggacat 3480
tcttgtaagg ctttgattag ttttctgtt gctttgttgc actgaaatgg aattcccatg 3540
cccctacccc ttacccagct tttttgaaca tggaaagaaa atttaataac tttttaaagt 3600
gacataattt acatgcaata tgttatcaa ctcaagaatt taatatagtt ggtacacaac 3660
tagttttgtt tataaattgg agatgcaaact agcaaaaacta aatacttgct ccatttaca 3720
actacttgat tttattgtac aagttgaaat atgctctttt gtttgggtta cagtatgctt 3780
gctctaagtc aaattccaag gaactaatct cttctcctgg agttgcattg attcagtatt 3840
acaaatatat agcacatcac ctgggacttg gcaatctt 3878

```

&lt;210&gt; 92

&lt;211&gt; 5674

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7376372CB1

&lt;400&gt; 92

```

gggagagcga gactccgagc gtgcgggcca ggaagctggg cggggggcgg agacgggggc 60
cgggacgacg gctggggcgc cgggggtcgcc agctcctcgc ccacctccct ccctgtggga 120
gaagaggagc aggggaagtg actgcggagt tgatgaactt tgcacatcca gaaacgccat 180
tttgattcct ttccggaac aagtttgcct ctctcctcgg tctctccctg aagcccaccg 240
gtccccaacg catcatgccg gggagccaag cgtcctgccc ggccgttgct ccgactgtgt 300
gttccaggag tgggtggctc gaggtgtgac cctgcccacg tttggggcca gccagggtca 360
gcccccaat aaggagggca gcttgataac acaaagaaaa cagtgtcaac gagtactacc 420
aagagaagaa gacaagaggt caacacagac tcatttttcta ctccgtgtga atgatagcta 480
cagcagggga aagtttcata gtctatcagt gggtcagaaa atggagtttt atagcagagg 540
cttcttagaa gcttaaaccc ctgtcccaat gacgtcaagt tcgcccgtcg gcttgggaagg 600
ttcagaccta tcttccatca acaccatgat gtcggcggtc atgagtgtag ggaagggtcac 660
agagaatggc gggagccccc aggggatcaa gtccccctcg aagcctccag gaccaaatcg 720
gattggcaga aggaaccagg aaacgaaaga ggagaagtct tcctataact gccccctgtg 780
tgagaagatt tgcactacc agcaccagct gaccatgcac attcgccagc acaacacaga 840
cactggagga gccgaccact catgcagcat ctgcggaag tcaactgagct cggccagctc 900
cctcgatcgc cacatgctgg tgcactctgg cgagaggcct tacaagtga ctgtgtgtgg 960
ccagtcattt accaccaatg ggaacatgca cagacatatg aagatccatg agaaggacct 1020
taacagtgcc acagccacag cccctccatc tcctctgaaa cgtaggcgat tgtcctccaa 1080
gaggaaactg agtcacgatg ccgagtcaga gagagaagac ccagcaccag ctaaaaagat 1140
ggtagaagac gggcagtcag gtgacttggg gaagaaagct gatgaagtct ttcactgccc 1200
agtatgtttt aaggagtgtt tttgcaagta tggactggag acccacatgg agaccattc 1260
agataaccga ctaagatgtg acatttgttg tgtcaccttt cgaacacatc gaggactgct 1320
gcgtcacaa cgccttgctc acaacaact tcccagggat gcaatgggca gacctttcat 1380
acagaacaac ccttcaattc ctgctggctt ccacgactta ggattcacgg acttctcctg 1440
taggaagttt cctcgcatct ctcaggcctg gtgcgaaaca aacctgcgga ggtgcatcag 1500
cgagcaacac cgttttgtct ggcacacctg tgacaaggcg tccccatgc tctgctcagt 1560
ggctctgcac aagcagaccc acgtggcggc agaccagggt caagaaaagc cgcaggccac 1620
gccccctgct ggtgacgccc tggaccagaa gggcttccctg gccttgettg gcctgcagca 1680
caccaaagac gtcaggcctg cccccgccga ggagcccctg ccggatgaca accaggcaat 1740
tcagctccag acactcaagt gtcagctacc tcaggacccc ggctgcacca acctgctgag 1800
cctgtcacct ttcgaagctg cttccctagg cggttctctc acagtctctc ccgcaaccaa 1860
ggacagcata aagcacctgt ccctgcagcc cttccagaag ggcttcatca tccagcctga 1920
cagcagcatt gtggtcaagc ccatctcttg cgagtcggcc atcgagctgg cagacatcca 1980
gcaaatctcg aagatggcag cctcggctcc ccctcagatc agtcttccgc ccttctccaa 2040
ggccccctgc gcccactgc aggcgatctt caagcacatg cccctctga aggcacaagc 2100
cctggtcaca ccacggacgg tgggtggcca cctccacgce cccgctctc atcaacgcca 2160
gcaggcttcc cggctgtatc agccccagcc tgccgccacc gcccctgaag ctccctcaa 2220
gtcactgga ggggctctcc aacgcccacc tgctgcagtc caagtccggg acccagcccc 2280
acgcggccac gggctctctc ctgcagcgca gccgcgggcg gagctgccgg gccagcctga 2340
gatgaagacg cagctggagc aggacagcat catcgaggcc ctgctgccgc tgagcatgga 2400
ggccaagatc aagcaggaga tcacagaggg ggaactcaag gccttcatga cagcgcccgg 2460
cggcaagaag acgcccgcga tgcgcaaggt gctctaccct tgccgcttct gcaaccaggt 2520
gtttgccttc tcgggggtct tgcgtgccc cgtgcgctcc cacctgggca tctcgccata 2580
ccagtgaac atctgcgact acatcgccgc cgacaaggcc gcgctcatcc gccacctgcg 2640
cacgcacagt ggggagcggc cctacatttg caagatctgc cactaccctt tcaactgtca 2700
agccaactgc gagcggcacc tgcgcaagaa gcacctcaag gccaccgca aggatatcga 2760
gaagaacatc gagtatgtga ttagcagcgc ggcgagctg gtggacgcct tctgcgcccc 2820
ggacaccgtg tgccggctgt gcggcgagga cctcaagcac tatcgtgcc tgcgcatcca 2880
catgccgacg cactgcggcc gcggcctggg cggggggccac aagggccgca agcccttcga 2940
gtgcaaggag tgcagcgccg cgttcgcggc caagcgcaac tgcattccacc acatcctcaa 3000
gcagcacctg cacgtgcccg agcaggacat cgagagctac gtgctggccg ccgacggcct 3060
gggccccgca gaggcgccgg ccgctgaggc gtcggggcgc ggggaggaca gtggctgcgc 3120
tgcccttggt gactgcaagc ccctcactgc cttcctggaa ccccaagaac gctttcttca 3180

```

```

caggggcccc acccagcctc cacctcccca tgtctcgatc aagttggagc ccgccagtag 3240
ctttgcggtg gacttcaatg agcccctgga cttctcgagc aagggcctgg ccctgggtcca 3300
agtgaagcag gaaaacatct cctttctgag cccttcttcc ctggtccctt atgactgctc 3360
catggagccc atcgacctgt ccatcccca gaacttcagg aaaggggaca aggatttggc 3420
cactcccagc gaagccaaga agcctgagga ggaggcgggg agcagcgagc agccctctcc 3480
ctgcccagca cccggccctt ctcttctgt aactttgggg ccagcgga tcctggaaag 3540
ccccatggcc cctgctccgg cggccacccc ggaaccccca gcacagccc tgcaggggcc 3600
tgttcagctg gcggtcccaa tctactctc agccctggtc agcagccctc cactcgtggg 3660
cagctcagcc ctcctgagtg gcacagcctt gctgctcca ctggggccca agccccgcct 3720
gcttttgcca aagcccccg tgacagaaga gctgcccccg ctggcctcca ttgcccagat 3780
catctcatct gtatcctcgg cccccaccct gctgaaaacc aaggtggcgg acccagggcc 3840
cgcaagcact ggcagtaaca ccacggcttc agacagctta ggaggttctg tccccaaagc 3900
cgccaccacc gccaccccg ctgccaccac cagcccaaaa gactctagtg agcctcccg 3960
tccagccaca ggcccagagg ctgctctcc caccgagcag ggcccagcgg gcacgtcgaa 4020
gaagaggggg cggaaaaggg ggatgaggag ccgaccccg gcacaacagc gcgggggtga 4080
cctgtactcc agcggggagt ttgccagcat cgagaagatg ctggccacta cagacaccaa 4140
caagttcagt ccgtttctgc agacagcggg ggacaacact caggatgagg tggccggagc 4200
ccctgccgac caccatgggc ccagtgatga agagcagggc agtccccag aagacaagct 4260
gtcggggggt aagcggaact cgtacaccaa ctgctgcag aagatcacct gtccccactg 4320
tccccgggtt ttcccttggg ccagctccct acagaggcac atgctcacac acactgacag 4380
tcagtcggat gcgggagact gcagccgcgg cgggcgaagt ggctatgacc tcacctcacg 4440
ggacagagag cagccgtcgg agggcgccac tgagctccgc caggtcgcag gggatgcgcc 4500
tgtggagcag gccacggcgg aaacggcctc gcggtgcacc gggaagagca cgggcgtggg 4560
gagagccatg agccggagga ggagcatggc actgaggaga gcactgggga cgccgacgcg 4620
gaagaggacg cgtcgagcaa ccagagcctg gacctggact tcgccaccaa gctcatggac 4680
ttcaagctgg cggaggggcga cgcggaggca ggccgggggc gcggcctcgc agggagcaga 4740
gctcgcctgc gacgcctgtg ggaagagctt caagttcctg ggcaccctga gccgccaccg 4800
gaaggcgcac ggccgcccag gaggcccaag gacgagaagg gagatggcgc cagcactgca 4860
gaggaggggc cccagcccg ccctgaacag gaggagaagc ccccgagac cccggcagag 4920
gtggtggagt cggcccggt cggggaggcc cggccggaa aagctcgcgg agggagcaga 4980
gggccccctc gacggggaga gcgcggccga gaaaaggtcc tcagagaaga gcgacgatga 5040
caagaaacca aagacagact ccccaaaaag cgtggccagc aaggcagaca agaggaagaa 5100
ggtctgcagc gtgtgcaaca agcgtttctg gtcgctgcag gacctgagcc ggcacatgcg 5160
ctccacaca ggggaaaggc catacaaatg tcagacctgc gagcgaacct tcaccttgaa 5220
gcacagcctg gttcgccacc agcggatcca ccagaaagcc aggcattgcca aacaccacgg 5280
gaaggacagc gacaaggaag agcggggtga ggaggacagc gagaatgagt ccaccacag 5340
cggaacaac gccgtctcag agaacgaggc tgagctggct cccaatgcca gcaaccacat 5400
ggctgtcacc cggagccgga aggagggtt ggccagtgcc accaaggact gcagccacag 5460
ggaggagaag gtcacgcagg gtggccgtct gagcctggcc agggtgacct taaccagag 5520
agccggcg ccctggggca ggacctgctg gagccgcgca gcaagaggcc tgccaccca 5580
atcctggcac agctgatggc gcctcccagc tcgtcgtagg ggatggagt acagcctcag 5640
ccccctcagc acagacaaaa gccagcagag caaa 5674

```

&lt;210&gt; 93

&lt;211&gt; 3188

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2754344CB1

&lt;400&gt; 93

```

cccgccatct ctatcagccc cagggatttg gcaaggagg taggcggctc cctttcggcc 60
actcattggt caaatcctaa tggaggtaag gcaactctgc agtctctatt ggtccctagg 120
ctccaaaaag gggggggaaa aacagaccga caggcctcta ggcgggaagc gcggagtgcg 180

```

```

ccggcgcgta gtcgggggacg ccggggccgag gctttgctag ggaacactgt tatcgccccg 240
ccccctcgga gctttttctc cagttaactg tcggagtagc gggggctccg gcgccgggcy 300
acatgccggt gcgcttcaag gggctgagtg aataccagag gaacttcctg tggaaaaagt 360
cttattttgtc agagtcttgt aattcctccg tggggcgaaa gtacccatgg gctggactta 420
gatcagatca attagggcatc acgaaagagc caagttttat ttcaaaaaga agagtccctt 480
accatgaccc acagatttca aaatctctgg agtggaatgg agctatctca gagagcaatg 540
tggttgcatc accagaacca gaagccccgg aaacaccaa atcacaagaa gcagaacaaa 600
aggatgttat tcaagaaaga gttcactcac tagaagcttc cagggttccc aaaagaacca 660
gatctcactc tgcagactcc agagctgaag gggcttcaga tgtggaaaat aatgagggtg 720
taacaaacca tacaccagtt aatgaaaatg tggaaactgga acattctacc aaggttcttt 780
cagaaaatgt agataatggg ttggatagac ttctgcgtaa gaaagctgga ttgactgttg 840
ttccttcata taatgccttg agaaattctg aatatcaaag gcagtttggt tggagactt 900
ctaaagaaac tgctccagct tttgcagcca atcaggtttt ccacaataaa agccagtttg 960
ttccaccatt caaaggtaac tcagtcatcc atgaaactga atacaaaaga aatttcaagg 1020
gtttatctcc agtgaaagaa ccaaaattaa gaaatgattt gagagaaaac agaaatcttg 1080
aaacagtgtc tcctgaaagg aagagtaata aaatagacga tcgtttaaaa ttggaagcag 1140
agatggaatt aaaagactta caccagccta aaaggaagct tactccttg aaacatcaaa 1200
ggcttgggaa ggtgaattcc gaatatagag caaaatttct gagcccagct cagtatttat 1260
ataaagctgg ggcttggaaca catgtaaagg gaaacatgcc aaatcagggt tctctaaatg 1320
ccatgtggta tgctgagggt aaagaactcc gagaaaaggc tgagttttat aggaagcgag 1380
ttcaggggac gcatttttct cgggaccatc tgaatcagat tttatctgat agcaactgct 1440
gttgggatgt ctctcaacc acaagctcag aaggaaccat tagtagcaac atcagagcat 1500
tagatcttgc tggagatcct acaagccata agactttgca gaaatgtcct tctacagaac 1560
cagaagaaaa aggaaatatc gtggaagaac agccccagaa aaataccacg gagaaatttg 1620
gtgtgtcagc tcccaccata cccgttagaa ggcggctggc ttgggataca gagaacacaa 1680
gtgaagacgt acagaaacag cccggggaga aagaggagga ggacgacaat gaagaggaag 1740
gggacaggaa aacgggcaag caggctttta tgggagagca agagaagttg gatgtacatg 1800
agaaatctaa ggcagataag atgaaagaag ggtcagatc ttctgtatcc tcagaaaaag 1860
gaggccggct tcctactccc aagctgagag aacttggtgg aatccagagg actcatcat 1920
atctcactat tccagtggtt ggtggtgctg ttttagtgtc tccatctaag atgaagcctc 1980
cagccctaga acagaggaaa agaattgacct cttaggattg tttagaaact tcaagaatg 2040
attttactaa gaaagaaagt cgtgctgtat ccctactgac ttctccagct gctggtataa 2100
aaacagttga tcctctgcct ttgcgggaag attctgaaga caatatccac aaatttgctg 2160
aggcaactct tccagtttca aaaattccaa aatacccaac aaatccccct ggacagttgc 2220
cttctccacc acatgttcca tcctactggc atccctctcg acgaattcag ggctctctta 2280
gagatccaga gtttcagcac aatgtgggaa aagcaaggat gaacaatttg cagttacctc 2340
aacatgaagc ctttaatgat gaagatgagg acagattgtc tgagatttct gctcgctctg 2400
cagcttctag tctccgggct tttcaaactc tggcacgagc taagaaaagg aaggagaatt 2460
tctggggtaa aacataaacc tagctgagtt gcctttatct aaggaaacat tggcataatt 2520
tttctttatt gcttactgtg ttaaaatttt ttgagtgaag ttttctctaa cattttctac 2580
tttgaatat ttaatttcca atagaacagt ttacttgga attgttaaca taggcttata 2640
caaatttaac ttttaaagaa ttctttttac atctatccca gagtacaaag acctagatta 2700
ttgaattctt cacatgtaac tttcgtaata ggtgtcatgg atagattttc acctatccag 2760
agagtattct aaaaaataa acaaagtga gtaaaagtta agatcccagt gtgaagtata 2820
tattccccta agatttgata tatgtatctt tataattgat gatgttgtaa aatgaatttg 2880
ttgttccctt aacctttgct gtagtgtttc tttttatttt agttccttct atacatgttt 2940
gcagtgttaa gataattgta ctttgaaaac accagctgta tatatttttc ttatttatgt 3000
aacaaagtgt gctgaaagat atgtatattt ttgatctttg agtgtattct atttgtgtaa 3060
ggactttggc ttacatttga ataattgtct gattttgaaa aattacctgt ataattggaga 3120
attaaaactt tgtagacatt ttgaaataaa aatggatcaa gtgcgaaaaa aaaaaaaaaa 3180
aaaaaaaa

```

&lt;210&gt; 94

&lt;211&gt; 2209

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens



&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 8268822CB1

&lt;400&gt; 94

```
agcgggttta ctccggcgcg cagtgccggc cggggcgggg atacgccgtg tgcgcggccg 60
ggggcggaac ggaggaggag gcggtgggtg cccggctgcg gggtaggagt ccgcggcagc 120
ctccgggtaa gccaaagcgc gcgcagtgct gagttccgc acgcgcgaga gccatggaga 180
tcggcaccga gatcagccgc aagatccgga gtgccattaa ggggaaatta caagaattag 240
gagcttatgt tgatgaagaa cttcctgatt acattatggt gatggtggcc aacaagaaaa 300
gtcaggacca aatgacagag gatctgtccc tgtttctagg gaacaacaca attcgattca 360
ccgtatggct tcatggtgta ttagataaac ttcgctctgt tacaactgaa cctcttagtc 420
tgaaagtctt tgataccaac atctttgata gtaacgtgcc ttcaaacaag agcaatttca 480
gtcggggaga tgagaggagg catgaagctg cagtgccacc acttgccatt cctagcgcg 540
gacctgaaaa aagagattcc agagtttcta caagttcgca ggagtcaaaa accacaaatg 600
tcagacagac ttacgatgat ggagctgcaa cccgactaat gtcaacagtg aaacctttga 660
gggagccagc accctctgaa gatgtgattg atattaagcc agaaccagat gatctcattg 720
acgaagacct caactttgtg caggagaatc ccttatctca gaaaaaacct acagtgcac 780
ttacatatgg ttcttctcgc ccttctattg aaatttatcg accacctgca agtagaaatg 840
cagatagtgg tgttcattta aacagggtgc aatttcaaca gcagcagaat agtattcatg 900
ctgccaaagc gcttgatatg cagagtagtt gggatatatg aacaggacgt ttgtgtgaac 960
cagaggtgct taacagctta gaagaaacgt atagtccgtt ctttagaaac aactcggaga 1020
aaatgagtat ggaggatgaa aactttcgga agagaaagtt gcctgtggta agttcagttg 1080
ttaaagtaaa aaaattcaat catgatggag aagaggagga agaagatgat gattacgggt 1140
ctcgaacagg aagcatctcc agcagtggtg ctgtgcctgc aaagcctgaa aggagacctt 1200
ctcttccacc ttctaaacaa gctaacaaga atctgatttt gaaggctata tctgaagctc 1260
aagaatccgt aacaaaaaca actaactact ctacagttcc acagaaacag acacttccag 1320
ttgctcccg aactcgaact tctcaagaag aattgctagc agaagtggc cagggacaaa 1380
gtaggacccc cagaataagt cccccatta aagaagagga acaaaaagga gattctgtag 1440
aaaaaaatca aggaatac aaagggcaat tattatccc actgcaaatc gaccagtaa 1500
tggcagaaac tctgcagatg agtcaagctg agatgagtga actgagtgtg gcacagaaac 1560
cagaaaaact tttggagcgc tgcaagtact ggcctgcttg taaaaatggg gatgagtgtg 1620
cctaccatca ccccatctca cctgcaaag ccttcccaa ttgtaaat tctgaaaaat 1680
gtttgtttgt tcacccaaat tgtaaatatg atgcaaagtg tactaaacca gattgtccct 1740
tactcatgt gtagtagaaga attccagtag tgtctccaaa accagcagtt gcaccaccag 1800
caccaccttc cagtagtcag ctctgccgtt acttccctgc ttgtaagaag atggaatgtc 1860
ccttctatca tccaaaacat tgtaggttta aactcaatg tacaagaccg gactgcacat 1920
tctaccatcc caccattaat gtcccaccac gacatgcctt gaaatggatt cgacctcaa 1980
ccagcgaata gcaccagtc ctgcctggca gaagatcatg cagtttggaa gttttcatgt 2040
actgatgaaa gatactctac agaacttgtc aaatctttga aacttggaa atattgcttt 2100
cataatatga agttttattg cctatctaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 2160
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaagaag ccaccccc 2209
```

&lt;210&gt; 95

&lt;211&gt; 2783

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1814553CB1

&lt;400&gt; 95

```
ggagcaggac gcgcgggggc cgcctcctcc cgcacggacc catgaaccag ccggcgggcg 60
cggcggtctc gcagggagcg cttgaagcgc agccagaaga gcaccaaggt ggagggccca 120
gagccagtgc cagccgaggc ctcgctgagt gccgagcaag gaacgatgac ggaggtgaag 180
```

```

gtgaagacag agctgcccga tgactacatc caggaggtga tctggcaggg cgaggccaag 240
gaggagaaga aggcgggtcag caaggatggg accagcgacg tgccctgccga gatctgcgtg 300
gtgatcggcg gcgtccgcaa ccagcagacc cttgggtctt acgaatgcgg aatctgtggc 360
aagaagtaca agtattacaa ctgcttcag acccagctgc gggcgaccg agacaccgaa 420
gccacctcag gggaggggagc ctcccaaagc aacaatttca ggtacacatg tgatatctgc 480
gggaagaagt acaataacta cagctgtttc caagagcacc gagacctgca cgcagtggat 540
gtgttttagtg tggaagggggc ccctgagaac cgggcagacc ccttcgacca aggtgtcgtg 600
gccacggacg aggtgaagga ggagccccg gagccattcc agaaaatcgg gccaaaaact 660
ggcaattaca cctgtgaatt ctgaggcaaa cagtacaagt actacactcc ctaccaggag 720
catgtggcct tacacgcccc catcagtact gccccgggt gggagccacc ggatgatcca 780
gacacgggct ctgagtgttc acatccagag gtctcccat ctccacgctt cgtggcagcg 840
aagaccaga cgaaccagtc ggggaaaaaa gctccggcct ccgtggctcg atgcgccacc 900
ctcttacacc gcaccctcc agccacccaa acccagacgt tccgcactcc aaattcggga 960
tctccggcga gcaaggcaac cgcagaaagc gctttcagtc ggagagtaga agggcaagca 1020
caaaaccact ttgaagagac gaacagcagt tcgcagaact ccagcgaacc ctacacctgc 1080
ggcgccctgt ggatccagtt ccagttctac aacaacctgt tggagcacat gcagtcccat 1140
gcagctgaca atgaaaacaa cattgcctcc aaccagtccc gatcgccacc tgctgttcta 1200
gaagagaagt ggaaacctca ggcccagagg aacagtgcc ataacaccac taccagtggg 1260
ttaacacca acagcatgat ccccgaaaag gagcggcaga acatcgcaga gcggctgttg 1320
agggctcatgt gtgccgacct ggggtgcactg agcgtggta gcgggaagga gttcctgaag 1380
ttggcccaga ccttagtaga cagtgggtgc cgctatgggg ccttctcggt cactgaaatc 1440
ctgggcaact tcaacacgct ggcgctgaag cacctgccac gcatgtacaa ccagggtgaag 1500
gtgaaagtga cctgtgcctt gggcagcaat gcctgcctag gcatcgggtg cacctgccac 1560
tcccagagtg ttggccctga ctctgtctac atctcacag cctaccaggc cgagggcaac 1620
cacatcaaga gctatgtgct tgggtgtgaag ggtgcggaca ttcgcgacag cggtgacctt 1680
gtgcaccact ggggtgcagaa cgtgctgtcg gagttcgtga tgtcggagat caggacagtg 1740
tacgtgacgg attgccgggt gagcacgtcc gccttctcca aggcgggcat gtgccttcgc 1800
tgctcagcct gtgccttgaa ctcggtgggt cagagcgtgc tgagcaagcg gacactgcag 1860
gccgcagca tgcacgaggt catcgagctg ctcaacgtgt gcgaggacct ggcggggtcc 1920
acgggcctgg ccaaggagac cttcgggtcg ctggaggaga cgtctccacc accctgctgg 1980
aactcgggtg cggactcact gttgctgggt catgagcgct atgagcagat ctgcgagttc 2040
tacagccggg ccaagaagat gaacctcatc cagagcctca acaagcacct gctcagcaac 2100
ctggcgggcca tctgacgcc ggtgaagcag gcagtcacg agctgagcaa cgagagccag 2160
cccaccctgc agctggtgct gcccacctac gtcaggctgg agaagctgtt cacggccaag 2220
gccaacgacg caggcactgt cagcaagctc tgccacctct tcctggaggc gctcaaggag 2280
aacttcaagg tgcaccggc ccacaagggt gccatgatcc tggaccgca gcagaagctg 2340
cggcctgtgc caccctacca gcacgaggag atcatcggca aggtctgtga gctcatcaac 2400
gaggtgaagg agtcctgggc cgaggaggcc gacttcgagc ccgctgcaa gaagccccgc 2460
tctgctgccg tcgagaacct cgcagctcag gaagatgatc ggctaggcaa aaatgaagtg 2520
tacgattacc tgcaggagcc cctcttccag gctaccctg atctcttcca gtactggctg 2580
tgcgttacct aaaagcacac aaaactcgcc aagctcgct tctggctcct ggcggttccg 2640
gccgtgggcy ccagaagcgg gtgtgtaaat atgtgtgaac aagcgcttct aatcaaacgy 2700
agggcggtgc tcagtccaga agatatgaat aaactcatgt ttctgaaatc caacatgctt 2760
taagacttga cttcggggga aaa 2783

```

&lt;210&gt; 96

&lt;211&gt; 1610

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 71217830CB1

&lt;400&gt; 96

aattaattat atgtgacacc actgcactcc agccttggtg acagagtga actccatcta 60

```

aaaacaaaac aaaatgaaac aaaaacccaaa taccctttaa atacatcctc gtgtactggg 120
gtgtatgtat ctatgggatg aattcctaaa agttgaccta tttttttaag ttccatctaa 180
ataaaatttc tcctttacaa catccactat tgtcatttct cctccccgtg gtttttgggt 240
cttttttagcc tcttgttcat ttactgcctt tcatgttccc atgctattgc cttgggtttt 300
taacaaaatt cttcctctgt gcctatcctg tcttacctac aaatagaatc ttcttgaggg 360
caaggatggg ttcttgtacg gtgtattggc actcggtaaa taaataatga gagagggatt 420
gagtgttggg atgagaaaag gagcctaata gttgcatttg gatgaatggg tgctgagaga 480
tgggtatgtac atttatacat tcttgtctca tcttgtaaaa tgtttgaaaa taagagtaat 540
tttcaacttg atttttctta ttttaggcaa gtatatcgcg tcaacacagc gacctgacgg 600
gacctggcgc aagcagcgga gggtgaaaga aggatatgtg ccccgaggag aggtcccagt 660
atatgaaaac aagtatgtga agtttttcaa gagtaaacca gagttgcccc caggggctaag 720
ccctgaggcc actgctcctg tcaccccatc caggcctgaa ggtggtgaac caggcctctc 780
caagacagcc aaacgtaacc tgaagcgaag ggagaagagg cggcagcagc aagagaaagg 840
agaggcagag gccttgagca ggactcctga taagggtgtc ctggaagaga cagcccaact 900
ccccagtgtc ccacagggtc ctcgggcagc cccacacagc gcacttgacc agcctgactc 960
agctgccacc actgagaaaag ccaagaagat aaagaaccta aagaagaaac tccggcaggt 1020
ggaagagctg cagcagcgga tccaggctgg ggaagtcagc cagcccagca aagagcagct 1080
agaaaagcta gcaaggagga gggcgctaga agaggagtta gaggacttgg agttaggcct 1140
ctgaggcctt tggggaatag ggaatggact gcagaacaaa ccgtgggggt ctctgggggtc 1200
tgggggaata cgggcaacag cagtcaggag ggggtacccc catactgggt tacttccacc 1260
tctgcggcc cagctctgtc ctccagagcc tagcgtctcc ctcaatcctt cccttttctt 1320
cccaacttct actttttgga ctttccccct cccattccca gtgttcaaaa tctcagtgc 1380
taccacaggt accttttctg ctgatttggg tgtcttgtt aaaagaaaat cagggtgggtg 1440
ggaatctctt ggagaactga ggctgagggt agagggagta tgcccaagtc ttggagtctt 1500
ggttcctgtt cgcggtgttt atgggttatt tccctctcca tccctcattt tttttttttt 1560
tttaaaaaaa gcaaaaatga gaataaacac aagtagactt gtaaacaaaa 1610

```

&lt;210&gt; 97

&lt;211&gt; 3467

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7506252CB1

&lt;400&gt; 97

```

gcgggcgccg gcgccatgga gcggctgcgg gacgtgcggg agcggctgca ggcgtgggag 60
cgcgctgttc gacggcagcg cgggcggcga ccgagccagg acgacgtgga ggcggcgccg 120
gaggagaccc gcgcgtctta ccgggagtag cgcactctga agcgtaccac gggccaggcc 180
ggcggcgggc tcgcagctc cgagtcgctc ccgcggcgcg ccgaagaggc gccagagccc 240
cgctgctggg ggccccatct gaatcgggct cggaccaaga gtccacagcc tacgccaggg 300
cggagccgcc agggctcggg gccggactac gggcagcggc tcaaggccaa tctgaaaggc 360
accctgcagg ccggaccagc cctgggcccgc agaccgtggc ctctagggaag agcctcatct 420
aaggcatcca ccccaaagcc ccaggtaca gggcctgtcc cctcctttgc agaaaaagtc 480
agtgatgagc ctccacagct ccctgagccc cagccaaggc caggccgggt ccagcatctg 540
caggcatccc tgagccagcg gctgggctcc ctatgcctg gctgggtaca gcgatgtcac 600
agtgaggtcc cagattttct gggggccccc aaagcctgca ggcctgatct aggctcagag 660
gaatcacaaac ttctgatccc tggtagagtc gctgtccttg gtcctggtgc tggctcccag 720
ggcccagagg cttcagcctt ccaagaagtc agcatccgtg tggggagccc ccagcccagc 780
agcagtggag gcgagaagcg gagatggaaac gaggagccct gggagagccc cgcacaggtc 840
cagcaggaga gcagccaagc tggaccccca tcggaggggg ctggggctgt agcagttgag 900
gaagaccctc cagggggaacc tgtacaggca agccctgcag cagcccactg 960
aacccagggt accacggact cagcccctcc agtcaagcta gggctgggaa ggctgagggc 1020
acagccccc tgacatctt ccctcggtg gcccgccatg acaggggcaa ttacgtacgg 1080
ctcaaatga agcagaaaca ctacgtgcgg ggccgggcac tccgtagcag gctcctccgc 1140

```

```

aagcaggcat ggaagcagaa gtggcggaag aaaggggagt gttttggggg tgggtggtgcc 1200
acagtcacaa ccaaggagtc ttgtttcctg aacgagcagt tcgatactg ggcagcccag 1260
tgtccccggc cagcaagtga ggaagacaca gatgctgttg ggccctgagcc actggttcct 1320
tcaccacaac ctgtacctga ggtgccagc ctggaccca cctgtctgcc actctactcc 1380
ctggggccct cagggcagtt ggcagagacg ccggctgagg tgttccaggc cctggagcag 1440
ctggggcacc aagcctttcg ccctgggcag gagcgtgcag tcatgcgcat cctgtctggc 1500
atctccacgc tgctgggtgt gcctacaggt gccggcaagt ccctgtgcta ccagctccca 1560
gcgctgctct acagccggcg cagccctgc ctacgttg tctgtctctcc cctgctgtca 1620
ctcatggatg accaggtgtc tggcctgcca cctgtctca aggcggcctg cataactcg 1680
ggcatgacca ggaagcaacg ggaatctgtc ctgcagaaga ttcgggcagc ccaggtacac 1740
gtgctgatgc tgacacctga ggcactggtg ggggcgggag gcctccctcc agccgcacag 1800
ctgcctccag ttgcttttgc ctgcattgat gaggccact gcctctccca gtggtccac 1860
aacttccggc cctgctacct gcgctgtgc aaggtgcttc gggagcgcat gggcgtgcac 1920
tgcttcctgg gcctcacagc cacagccaca cgccgactg ccagtgcagt ggcacagcac 1980
ctggctgtgg ctgaagagcc tgacctccac gggccagccc cagttccac caacctgcac 2040
ctttccgtgt ccatggacag ggacacagac caggcactgt tgacgtgct gcaaggcaaa 2100
cgttttcaaa acctcgattc cattatcatt tactgcaacc ggcgcgagga cacagagcgg 2160
atcgctgcgc tctccgaac ctgcctgcac gcagcctggg tcccagggtc tggaggctcg 2220
gcccccaaaa ccacagccga ggcctaccac gcgggcatgt gcagccggga acggcgggcg 2280
ccccaggcgg aagacctgcg agagctgcgc agacatgtgc acgccgacag cagggacttc 2340
ctggctgtga agaggctggt acagcgcgtg tcccagcct gcacctgcac ctgcaccagg 2400
ccgccctcgg agcaggaagg ggcctgggt ggggagaggc ctgtgcccac gtacccccct 2460
caagaggctg agcagcttag ccaccaagca gcccaggac ccagaagggt ctgcatgggc 2520
catgagcggg cactcccaat acagcttacc gtacaggctt tggacatgcc ggaggaggcc 2580
atcgagactt tgctgtgcta cctggagctg caccacacc actggctgga gctgctggcg 2640
accacctata ccattgccc tctgaactgc cctggggggc ctgccagct ccaggccctg 2700
gcccacaggt gtcccccttt ggctgtgtgc ttggcccagc agctgcctga ggaccaggg 2760
caaggcagca gctccgtgga gtttgacatg gtcaagctgg tggactccat gggctgggag 2820
ctggcctctg tggcggggc tctctgccag ctgcagtggg accacgagcc caggacaggt 2880
gtgcggcgtg ggacaggggt gcttgtggag ttcagtgagc tggccttcca ccttcgcagc 2940
ccggggggacc tgaccgctga ggagaaggac cagatatgtg acttcctcta tggcctgtg 3000
caggcccggg agcgccaggc cctggcccgt ctgcgcagaa ccttccaggc ctttcacagc 3060
gtagccttec ccagctgcgg gccctgcctg gagcagcagg atgaggagcg cagcaccagg 3120
ctcaaggacc tgctcgcccg ctactttgag gaagaggaag ggcaggagcc gggaggcatg 3180
gaggacgcac agggcccccga gccagggcag gccagactcc aggattggga ggaccaggtc 3240
cgctgcgaca tccgccagt cctgtccctg aggccagagg agaagtcttc cagcagggct 3300
gtggcccgcga tcttccacgg catcggaagc ccctgctacc cggcccagggt gtacgggcag 3360
gaccgacgct tctggagaaa atacctgcac ctgagcttcc atgccctggt gggcctggcc 3420
acggaagagc tctgcaggt ggcccgtga ctgcactgca ttggggg 3467

```

&lt;210&gt; 98

&lt;211&gt; 2392

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2270608CB1

&lt;400&gt; 98

```

actaggctga acggggccgag cactcgagcg gggacgcgca gctctgcaga gccccggggc 60
ggggcctcgg tccagggtgc cggtagcgct tcgcgaggcg aggaagtgtg tccctagcgg 120
cgccccgtgc agcgctcccg cgagacgctc acctgcgcc caggtagcgc gcgagggggc 180
gggggagggg ctgagccgga ggcggtcac ctggcgggac aggtgcctgg ctgctacaaa 240
ccatgcaatg agccatgccc cgccctggac acccccgcc agcatctggg cctccacgct 300
tgggaccgtg ggagcgggca acagagctat gtctggagac atatgataaa ccacctcagc 360

```

```

ccccaccaag cgcgcgcacc cgtagaccag accccaagga ccctggccac catgggccag 420
agagcattac cttcatctct ggctctgtct agccggccct tgagtcccc acctgtgtcc 480
tgctctggcg accctgggtg tgggagtggg gccgggctgc cttctgtctc cgccgctgcc 540
gggattgcct ccagcgctgt ggagcctgtg tgcggggatg cagccccctgc ctgtctactg 600
aggactccac tgaggggact gctgaagcca actggggcaa ggagcacaat ggagtgtccc 660
ccagccctga tcgtgcaccc ccagccggc gggatggcca gcggctcaag tcaaccatgg 720
gcagcagctt cagctacccc gatgttaagc tcaaaggcat ccctgtgtat ccctaccgga 780
ggggcacctc ccagccccc gatgoggact cctgtgtcaa ggagccactg gccgatcccc 840
caccatgcg acacagcctg ccagcacct ttgccagtag tcctcgtggc tccgaggagt 900
actattcttt ccatgagtcg gacctggacc tgcgggagat gggcagtggc tccatgtcga 960
gccgagaaat tgatgtgtct atcttcaaga agctgacaga gctgttcagc gtacaccaga 1020
tcgatgagct ggccaagtgc acatcagaca ctgtgttcct ggagaagacc agtaagatct 1080
cggaccttat cagcagcatc acacaggact accacctgga tgagcaggat gctgagggcc 1140
gcctggtagc cggcatcatt cgcattagta cccgaaagag ccgtgtctgc ccacagacct 1200
cggagggtcg ttcaactcgg gctgtgtccc caaccgtgc tgcacctgac agtggccatg 1260
agaccatggt gggctcaggt ctgagccagg atgagctgac agtgcagatc tcccaggaga 1320
cgactgcaga tgccatcgcc cggaaagctga ggccttatgg agctccaggg taccagcaa 1380
gccatgactc atccttccag ggcaccgaca cagactcgtc gggggcacc ttgctccagg 1440
tgtactgcta acccctgccca ggcccagctg ccacaccctt tctgggagaa gcatggccta 1500
cagaatgaag agggggacca ggaacccctg tgggagaggc ttagacctga agcagtgccc 1560
actctggctc ctcctgcctt ggctgactgg gttcctggac catgtgcatt tcaactggcc 1620
atgggatcta catctccttg catccccagc tggctgtatc cctgccaggg ccccttcctt 1680
cctgtctcat gtcttcagggt ggctgatca tggaaagtaa ggagttaggc attaccttct 1740
gggagtgaac cctgactcca tccccctatt gccaccctaa ccaatcatgc aaacttctcc 1800
ctccctgggg taattcaaca gttaaagaa gcttatctta aatgtattgt attgggggt 1860
gggcagggcc cactctatgt tatgttaagg agttggttct ggttcttggc tgatgttctg 1920
tatcttaaca tgaccacagt ttgtaagtac aaaggtaatt ggtatctgcc tgctttcttg 1980
gctgtccttt ctctcccatc tgcctcactt aactgctgcc cagagcagga gacaggatat 2040
gtgtccagcc tggcttagtg cccgcagtta aagtaggaag gataaaatgg gagtaatgct 2100
gccccattt caggttcaact gtgagttcca ggagataaag aaccaagcac agagtctgtg 2160
cctgagaaag gttggataaa tggtttctga cattgtttga gaggttcttc tagggagggc 2220
tgtccgaact gcctctgtc tcccagtcca ctgcctgcag gctggcctgc ttctgcccag 2280
ccacagttt ccccaaagcc ttgatcattt gttggcccta cccagccttg gacctgtgtc 2340
tccccctcca ctctctctgt ggtagtaaca gagcactgtg gttttttttt tt 2392

```

&lt;210&gt; 99

&lt;211&gt; 1326

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7502428CB1

&lt;400&gt; 99

```

tgtgtccct taagcatgtc tgttatgagc ttccttgact tcacacctct cagatttgac 60
aaacatgagg agaaacaaag caaactctct tttctgtggg ttaatttgct tctcattttg 120
ggctcctcgt atatcagagt gtgaatgata ctttattgtc ctcaactgtg cttcaattcg 180
tggctgtttt gcaggaggag ctccataaaa aaatgcagtc tttatgggaa aaagcttggt 240
aaaatctcag aaatctgaac atggaaacca caagaaccag atgctggaag gattatgtga 300
gtttaaggat agaagcaatc agagctgaat atcagaagat gcctgcattt ctccatgaag 360
aagagcaaca tcaactggaa aggctgcgaa aggagggcga ggacattttt cagcaactca 420
atgaaagcaa agccagaatg gaacattcca gggagctttt aagaggaatg tatgaggatc 480
tgaagcaaat gtgccataaa gcagatgtgg agctactcca ggcttttggg gacatattac 540
acaggatatg gtctctgtct ctgcaagtgt ctgagcctgt gaatccagag ctcaagtgcag 600
ggcccatcac tggactgctg gacagcctca gtggattcag agttgatttt actctgcagc 660

```

```

ctgaaagagc caatagtcac atcttcctgt gtggagattt gagaagcatg aatgttggat 720
gtgaccctca agatgatccc gatatacttg gaaaatctga atgttttctt gtatgggggg 780
ctcaggcttt cactctgggc aaatattatt gggagggtca tatgggggac tcttgggaatt 840
gggcttttgg tgtctgtaac aattatttga aagagaagag acagaatgac aagatagatg 900
gagaggaggg actctttctt cttggatgtg ttaaggagga cactcactgc agtctcttta 960
ccaccacccc acttgtgggtg caatatgttc caagacctac cagcacagta ggattattcc 1020
tggattgtga aggtagaatc atgagctttg ttgatgttga tcaaagtctt ctgatataca 1080
tcataccctaa ttgctccttc tcacctcttc tcaggcctat cttttgctgt agtcacttct 1140
gaccagagaa aagtcagaaa tgtgcctgta tgctctggga acctgtttat cccagaaaagc 1200
cctctttttc gcacctcatc aaacagaaca aataagttat atttaatgtc tttagttgca 1260
ttctaagtgc atcaaaactc atttatagtg tttctattaa atatggtgaa aacattaaaa 1320
aaaaaa                                     1326

```

&lt;210&gt; 100

&lt;211&gt; 2245

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 368741CB1

&lt;400&gt; 100

```

ggaccattct ggtcaggtgt ggggggttgt tgtgtacact ggtggccatt ttttggtttt 60
cagacaatgg tgagaatata tagggagctg ctgatgcaca ttgaggtctc ttgcttttagg 120
aactgatgac gcttgataat gtggctgggg acttcaggga agagtgggtt acctggacac 180
tgcttagaga atcctctcca ggaatgccac ccagcacagt tagaagaatg ggctctcaaa 240
ggaatttcca ggcctagtgt aatctcccag ccggagcaga aagaagagcc atgggtccta 300
ccactccaaa actttgagggc gaggaagatc ccgagggaaa gccacacaga ctgtgagcat 360
caggtggcaa agctcaatca ggacaattct gaaacagcag aacaatgtgg aacatcctca 420
gaaaggacca ataaagatct ttctcatact cttagtgtgg gaggaactg ggagcaaggc 480
ctagaattag aagggcaata tggaaacctt ccaggagagg gccagctgga gtccttttca 540
caggagaggg atttaaacaa gtcctgggat ggatatgtag gagagaagcc tatgtgtgca 600
gaatgcggga aaagctttaa ccagagttcc tatctcataa gacacctaag aaccacact 660
ggcgagaggg cctatactgt cattgagtgt gggaaaggct tcaaacagag ctgagacctt 720
gtcaccatc gcagaacaca cacaggagag aagccctacc aatgcaaggg gtgtgagaag 780
aaattcagcg acagctcaac actcatcaaa catcagagaa cccacacagg ggagagacct 840
tatgagtggc cagagtgtgg aaagactttt gggcggaagc cacacctcat aatgcaccaa 900
agaaccaca caggcgagaa gccctacgcg tgcctggaat gtcacaaaag cttcagtcga 960
agctcaaatt tcactactca ccagaggacc cacacagggg tgaagcctta caggtgtaat 1020
gactgtgggg agagttttag ccagagctcg gatttgatta agcaccaacg aaccacacg 1080
ggagaacggc ccttcaaatg cccggagtgc ggggaagggt tcaagatag ttctcatttt 1140
gtagctcaca tgagcactca ttcaggagag aggcctttca gttgtcctga ctgccacaaa 1200
agcttcagtc agagctcaca ttgggtcacg caccaaagaa cacacacagg tgagagacct 1260
tttaagtgcg aaaactgtgg gaaaggatc gccgacagct ccgccctcat taagcaccaa 1320
cgaatccaca ccggagaaag accctacaaa tgtggagagt gtgggaagag cttcaatcag 1380
agctcccact ttattacca tcagcgaatc cacttaggag acaggcccta tcgatgtcct 1440
gagtgtggca agaccttcaa tcagcgttcc catttcctca cacaccagag aacgcataca 1500
ggagaaaaac ctttccactg tagtaaatgt aacaagagct tccgtcagaa agcgcatctt 1560
ttatgccatc aaaacacca tttagtttag gaagtagtct ttggtgttca gctgctccct 1620
tgcacatttt cattgctact gtcttcaagc accccaaata gagaaaacct gggcgtcagt 1680
ggctcaattt gggcctgat ctattctccc tctttcttgt ctatgttata acagagagga 1740
taaaacttaaa ggggtccaat aacggtccga atacaaaagg cattccttca gtgtgtgact 1800
gactcttagg gaaatgtgag tttaatagtt gatgcccggc aggcgtgggt gctcaccctt 1860
gtaatccag cacttttggg aggccaaagg ggggtggatca cttgaggtca ggagttgaga 1920
ccagcctggt gagcatggtg aaacctcatc tctactaaaa atgcaaaaat tagctgggca 1980

```

```
tggtggcacg tgcctgtagt cccagctact cgggaggccg aggcagaaga atcatttgaa 2040
ctcagaaggc gcaggttgca gtgagttgag atcatgccac tgcactccag cctgggcaac 2100
agagagagac tctgtctcca agaaaattaa aaagttgatg cctagttact aaatgagaag 2160
tgagaaaaat gtggcctaga gatcacatgt tcaccaccta gtacagtgcc tggcacaaca 2220
tagatggtca ataactaaag tttgg                                     2245
```

&lt;210&gt; 101

&lt;211&gt; 547

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7506379CB1

&lt;400&gt; 101

```
gtggcggcga acgcggagag cacgccatga aggcctcggg cacgctacga gagtacaagt 60
taaagaagat gaagaagtct tcaggggaga ttgtctactg tgggcagggtg tttgagaagt 120
ccccctgcy ggtgaagaac ttccgggatct ggctgcgcta tgactcccg agcggcaccc 180
acaacatgta cgggaataac cgggacctga ccaccgcagg cgctgtcacc cagtgtctacc 240
gagacatggg tgcccggcac cgcgcccag cccactccat tcagatcatg aagggtggagg 300
agatcgcggc cagcaagtgc cgcgggccgg ctgtcaagca gttccacgac tccaagatca 360
agttcccgtc gccccaccgg gtcctgcgcc gtcagcacia gccacgcttc accaccaaga 420
ggcccaacac cttcttctag gtgcagggcc ctcgtccggg tgcagggccc tcgtccgggt 480
gtgccccaaa taaactcagg aacgccccgg tgaaaaaaaa aaaaaaaaaa aaaaaaaaaa 540
aattggt                                     547
```

&lt;210&gt; 102

&lt;211&gt; 880

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7506253CB1

&lt;400&gt; 102

```
tcggggccca aggctgaggg acccggtgga gcggaagtca ctccctgagg cagtggcgac 60
agcggcgcg agaggatgaa caacaagttc gacgctttga aagatgatga cagtggggac 120
catgatcaga atgaagaaaa cagcacacag aaagatgggtg agaaggaaaa aacggaacga 180
gacaagaatc agagcagtag caagagaaag gtggagcagt tctggagggtt ttatagccac 240
atggtacgtc ctggggaccc gacaggccac agtgacttcc atctcttcaa agaaggaatt 300
aaacccatgt gggaggatga tgcaataaaa aatgggtggca agtggattat tcggctgcgg 360
aagggccttg cctcccgttg ctgggagaat ctcathttgg ccatgctggg ggaacagttc 420
atggttgggg aggagatctg tggggctgtg gtgtctgtcc gctttcagga agacattatt 480
tcaatatgga ataagactgc cagtgaccaa gcaaccacag cccgaatccg ggacacactt 540
cggcgagtgc ttaacctacc tccaacacc attatggaat acaaaactca caccgacagc 600
atcaaaatgc caggcaggct gggcccccac aggcctcctt ttcaaaacct ctggaagccg 660
cggttgaatg tgccatgacc ctctccctct ctggatggca ccatcattga agctggcgtc 720
atcggagtct cttgttctgt tggcggtgta cctggaagat ccttctgtcc tggacaagag 780
gaattggaag agcattttat gttttaagaa caggctgaca cgcagcagct acaacaacag 840
ctgagatcac ttaataaatg gtgctaaact aaaaaaaaaa                                     880
```

&lt;210&gt; 103

&lt;211&gt; 813

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7506353CB1

&lt;400&gt; 103

```

cccggaaacc ccaaggccca gacggctctc agatccgggg gctgcggata aatccccgta 60
ggccgcgggc agcgagatgt tgcgttccgg tgtgggtgtg ggtgtgcctc cgacggcgtc 120
tcgggtgccag tgtcgagggt ctttctgctt agctaccggg agccgactac ggaggaggac 180
acctgagttt acgtctcttc catctgctgc tcgcctcagc tgcctgggtc cccgacgaga 240
gccagggtgac acttaactcc gccatctgcg ttttgagcac tgttctcata atggagtttc 300
ctgatttggg gaagcattgt tcagaaaaga cttgcaagca gctagatttt cttccagtaa 360
aatgtgatgc atgtaaaca gatctctgta aagatcattt tccatacgtc gcacataagt 420
gtccgttttc attccagaag gatgttcacg tcccagtatg cccactctgt aataccccca 480
tcccagtaaa aaagggccag ataccagacg tgggtggttg tgatcacatt gacagagact 540
gtgactctca ccctgggaag aagaaagaga agcacagaca ccctttggac cacagctgca 600
gacacgggag tcgccccacc atcaaagctg ggtgagaaga gactccgctg cgatggctcg 660
gagcagcgag tagcatgctg ggaagcagct tacactctag tgggaagtgg agccccattg 720
agcaccatcc cacactggct gctgatcttg tttgttgagg gagatgggac tataataaaa 780
ttgaatgctg aagttgaaaa aaaaaaaaaa agg 813

```

&lt;210&gt; 104

&lt;211&gt; 3014

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7506372CB1

&lt;400&gt; 104

```

cgctatggcg gcggtgttga ggttgggtac gggatgcggg gtctttgact gaaggggtag 60
gccaaagtga ggtatcaggg acgtcgcgcg gcacagaaga ggaccagcct ggacgcccgg 120
gacgtgttca tgtacggcgc gagcgggggc cgcgccaaac ccgagaggaa aagcggcgcg 180
aaggaggagg ccgggccagg cggtgccggc ggtgggggca gccgagtga gctcttggtt 240
ttcggctatg cctgcaagct gttccgggac gacgagcggg ccctggctca ggaacaggga 300
cagcacctca tcccctggat gggggaccac aagatcctca tcgacagata tgatggacgt 360
ggtcacctgc atgaccttct tgagtacgat gctgagtatt ccacgtggaa cagagattat 420
cagctgtctg aagaggaggc gcgaatagag gccctgtgtg atgaagagag gtatttagcc 480
ttgcatacgg acttgcttga ggaggaggca aggcaagagg aagaatacaa gcgattgagt 540
gaagcactag cagaggatgg gagctacaat gccgtggggt tcaactacgg tagcgactat 600
tacgacccgt cagagccgac ggaggaggag gagccttcca aacagagaga aaaaaatgag 660
gccgaaaatt tagaggaaaa tgaagagccc ttcgttgccc ccttaggatt gagcgtccc 720
tctgacgtgg agttgccacc aaccgctaaa atgcacgcca tcatcgagcg cacggccagc 780
ttcgtgtgca ggcagggagc acagtttgag atcatgctga aggccaagca ggcccggaac 840
tcccagtttg actttctgcg cttcgaccac tacctcaacc cctactataa gttcatccag 900
aaagccatga aagagggacg ctacactgtc ctggcagaaa acaaaagtga cgagaaaaaa 960
aatcaggag tcagctctga caatgaagat gatgatgatg aagaagatgg gaattacctt 1020
catccctctc tctttgcctc caagaagtgt aaccgccttg aagagctgat gaagcccttg 1080
aaggtagtgg acccagatca tcccctcgca gcacttgctc gtaaggcaca ggctgacagt 1140
tccactccca cccacacaa cgacagcggg gcgcctgtgc agccctccca ggtggagtag 1200
acggcagact cgaccgtggc agccatgtat tacagctact acatgctacc ggacggcact 1260
tactgcctgg cgccgcccc tcccggaatc gacgtgacta cttactacag cacccttctc 1320
gctggcgtga ccgtgtctaa ctcccctgga gtgacgacca ccgccccacc acctcctggg 1380
accacaccac caccgcccc aaccacagca gagactagca gcggggccac ctccacaacc 1440

```



```

accaccacaa gtgcacttgc ccccggtggcc gccatcatcc ccccgccccc cgacgtccag 1500
cccggtgattg acaagctggc cgagtatgtc gccaggaacg gcctgaagtt cgagaccagt 1560
gttcgtgcca agaattgatca aagatttgag ttcctgcagc cgtggcacca gtataatgct 1620
tattatgagt ttaagaagca gttcttcctc cagaaagaag ggggcgatag catgcaggct 1680
gtgtctgcac cagaagaggc tcccacagac tctgtctccg agaagccaag tgatgctggg 1740
gaggatggcg cgcctgaaga cgcagccgag gtgggagcac gggcaggctc aggcgggaag 1800
aaggaggcat cgtccagtaa gaccgtcccg gacgggaagc tggtgaaagc aaagcaaaag 1860
ctggaagatc gcctcgagc tgctgcccgg gaaaagctgg cccaggcgtc taaggagtca 1920
aaagagaaac agcttcaagc agaacgtaaa aggaaagcgg cgttattttt acagaccctc 1980
aaaaatcctc tgccggaagc agaagctggg aaaattgagg agagtccctt cagtgtcgag 2040
gaatccagca ctacgccctg ccctctactg actggaggca ggcctctgcc tacttttagaa 2100
gttaaaccac ccgataggcc ttcgagcaaa agcaaagatc caccgagaga agaagagaaa 2160
gaaaagaaaa agaaaaagca caaaaaaaga tctcgaacaa gatcacgttc tcccaagtac 2220
cattcgtcat ccaagtccag gtctagatca cactcaaaag caaagcattc tcttcccagt 2280
gcctatcgga cagtgcggcg gtccgaggtcc cgctcccggg cccctcggag gagagccac 2340
tcccctgaga gacggaggga agagaggagt gtgccactg cctaccgctg gagccgcagc 2400
cctggggcca gcaggaagcg gacccgctcc agaagtcccc acgagaagaa gaagaagagg 2460
cgggtcccggt cgcggaacca gtccaaggcc aggtctcagt cgggtgcacc cagcaagcag 2520
gcagcgcccc ggccccggcg ccccgcgccc cactcgcgcc actcagccag cgtctcccct 2580
gtggagagtc gggggtccag ccaggagcgc tccaggggag tctctcagga aaaagaagcc 2640
cagatctctt cagcaatcgt ttcttcctg cagagcaaaa tctctcagga tctcatggcc 2700
aaagtccagc cgatgcttgc agcttccaaa aacctgcaaa ccagcgcttc ctgagacggg 2760
gccagcggag gcagagcccg gaggtctgct gggcttctgg gcaggctcac gcagacgccg 2820
gccacaccat ccacctggcc gcctccatgg acccttggtg gcttttgtaa attaatTTTT 2880
gatgacattt tgagttttaa gatctctgac cagcagcttc ttacctgtat atttgtaaAT 2940
atatcatgtt tctgtgaaaa tgtattatga aataaaatgg gaggaaacac cttttctagc 3000
tagaaaaaaa aaaa                                     3014

```

<210> 105

<211> 772

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 7506335CB1

<400> 105

```

gccgggaccc gccctagggg ctccggtcgc cgcgcgctcg ccgctcgccc gccagcccg 60
ccgtgggtccg tggcggcgcg ctccacccgg cagggggagg cgcggggcgc accatggccg 120
cagacacgcc ggggaaaccg agcgctcgc cgatggcagg agcgccggcc agcgccagcc 180
ggaccccaga caagccccgg agcgcgcccg agcaccgcaa gagctcccgc cactcgaagc 240
tgagagaaggc ggacatcctg gagatgaccg tgagacacct ggggagcctg cgtcgcgtgc 300
aggtagcggc cgcgctcagc gccgaccccg ccgttctggg caagtaccgc gccggcttcc 360
acgagtgtct gggcgagggtg aaccgcttcc tggccggctg cgagggcgtc ccggccgacg 420
tgcgctcccc cctgtggggc cacctggcag cctgcctgcg ccagctggga ccctcccggc 480
gcccggectc gctgtccccg gctgcccccg cagaggcccc agcgcccagc gtctacgcgg 540
gccgcccgtc gctgccatcg ctcgggggcc ccttccctct gctcgcgccg ccgctgctgc 600
cgggtctgac ccgggcgctg cccgccgccc ccagggcggg gccgcagggc ccgggtgggc 660
cctggaggcc gtggctgcgc tgaggctgtg gcctgagact gactcggat taccggcgcc 720
ccgttctagg gccgtgccct taaccgagag gtagcagaga tactattctc tc 772

```

<210> 106

<211> 4206

<212> DNA

<213> Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 5546982CB1

&lt;400&gt; 106

```

cgcccaggtc tctgcgcggc gccgttccca gccgcagtc gcaggagccg ctgctgcagc 60
cggcagctgg gcgggattcc cggggtagtc gctgggagca aagctcgctt ctcccgcgcc 120
tgtgcttttg cctgccccgg ctacttctc cccgcgtcgg tccgctggcc acgtacagct 180
gaccagctca tcttcagct tcggaccgag gctccacggc cgccgcaggg acccgccac 240
cagaccctgc gctcggtagc cgggtgtgagt ggaacgtgcc gctgatgaca cctggcagaa 300
ggcagccatg gagggcgatg gctcagattc gccggttacc attaagaata tcgaaaggga 360
gctcatttgc ccagcatgca aggagctgtt taccaccca ttgattctc cttgccaaca 420
tagtatctgt cataaatgtg taaaagaact cctgctgact ctcgatgatt cattcaacga 480
tgtgggatca gacaactcca atcaaagcag tcctcgactt cggctcccct cccctagtat 540
ggataaaaatt gaccgaatta acagaccagg ctggaagcgc aattcattga ccccgaggac 600
aactgttttc ccttgccctg gctgtgagca tgatgtggat cttggagaac gaggaatcaa 660
tggtctgttt cgaaacttca ctttggaac tattgtggaa agatatcgct aagcagctag 720
ggcagccaca gccattatgt gtgaccttg taaaccacca cctcaagaat ccacaaaaag 780
ctgcattggc tgtagtgcaa gttactgcaa tgaatgctt aaaattcatc accctgggg 840
tactataaaa gctcaacatg agtatgttgg tccaactact aacttcagac ccaagatttt 900
aatgtgcccc gaacatgaaa cagagagaat aaacatgtac tgtgaattat gtaggaggcc 960
agtttgccat ctgtgtaagt tgggtggtaa tcatgccaac caccgtgtaa ccactatgag 1020
cagtgcctac aaaaccttaa aggaaaagct ttcaaaggat attgattacc ttattggtaa 1080
ggaaagccag gtgaagagtc aaatatctga actaaacttg ttaatgaaag aaacagagtg 1140
taatggagag agggctaaag aagaagcaat tacacatttt gaaaagctct ttgaagtctt 1200
ggaagagagg aaatcatctg ttttgaaagc aattgactcc tctaagaaac taagattaga 1260
caaatttcag actcaaattg aagagtacca gggacttcta gagaacaatg gacttgtgtg 1320
atatgctcaa gaagtgtcaa aggagacaga tcagtcttgc tttgtgcaga cagcaaagca 1380
gtccacctc agaatacaga aagccacaga atctttgagg agctttagac ctgcagctca 1440
gacttctttt gaagactatg ttgttaatac ctctaacaa acagaacttc ttggagaatt 1500
atcctttttt tctagtggca tagacgtgcc agagatcaat gaggaacaga gcaaagttaa 1560
taacaatgcc ttgataaatt ggaccatcc agaaaaggat aaagctgata gctatgttct 1620
tgaatatcgg aaaatcaata gagatgatga aatgtcatgg aatgagatag aagtgtgtgg 1680
aacaagtaaa ataattcaag acttgaaaa cagtagtacc tatgctttca gagtaagagc 1740
ttacaaggg tcaatctgta gtccttgag cagagaattg attcttcata ctctccagc 1800
tccagttttc agcttctctt ttgatgaaaa atgtggctat aataatgaac acctcctgct 1860
gaacttgaag agagaccgtg tagagagtag agctggattt aatcttctgc ttgctgcaga 1920
acgcatccaa gtgggttatt acacaagctt agactacatc attggagata ctggcattac 1980
aaaaggaaaa cacttctggg ccttccgtgt ggaaccatat tcatacttg taaaagtggg 2040
agttgcttct agcgataaac tacaagaatg gctccgttct ccccgggatg cagttagtcc 2100
aagatatgag caagacagtg ggcagacag tggaagttag gatgcctgtt ttgattcttc 2160
acaaccattt accttagtta ctataggcat gcagaaattt ttataacca agtcacctac 2220
ttcttcta at gaacctgaaa atagagttct ccctatgcca acaagtattg ggattttcct 2280
tgactgtgat aaaggcaaag taaatttcta tgatatggat cagatgaaat gcctttatga 2340
acgccaagtg gactgttcac atacactgta tccagcattt gcattaatgg gcagtggagg 2400
aattcagctt gaagaaccca tcacagcaaa atatctggaa taccaagagg acatgtagtt 2460
gaaacatctg atgtgactca ggataaaaa tgtaaaaaa gcaatgcctt agcgtaaac 2520
ttcgtaacta attacatata gctgagtggt ctgtcaccct acttggtgtt gagtatgatt 2580
cttcttgaaa tacagaaaat ctaaacagcc ttgtgtcttc atgtcgttgt tctgccttcc 2640
acaaatgatg aaagtgaaga tgttctttcc aagttaatac tgaagtgggt aactttgcaa 2700
ctggactgtc ttttacttta tgtacaggat atctaattta tttctgtcag ctggtgttgc 2760
tgacactttc cttatggtgg gggggatata aaagtattta ttatgattct cttctgtttg 2820
acttttctt tttatatcat gtttatgtcc taagagaaga ttttttttaa atgccttatg 2880
gataggaatt agtactaagc aatgtgtctt gcatttgttt tcaataggaa aacaaagttc 2940
tagatagggt aaaagaaaa caccataaaa gtggtagtat atttaagtg aaaattatgt 3000
attttaaggt gtgtaactat taaaaatata ttgcctggaa aaggaatact actgttttgt 3060

```

```

tttggttaaga ccacaaactg tattgctaag gacattttaaa gagtctgctt agccataaca 3120
ctcattttaac agcatgggta gtaagcatgg tttcttgtag aaaacacaaa gttctcagct 3180
aacacattaa acttctcact ttatatactt tttttcagag aaattttatct ttagtttttc 3240
ttattctgtg aaaatatttg ttcttatgtc aaagaattat ttttcagata tattggtaaa 3300
gatggactga tattaanaagt catatttatt cttttacccc cttactgtgt tttgtgattt 3360
tcatttttgg cccaattatt acatattcagg ttttcctgat ttatatttac atgtttgaaa 3420
agtttcattat atatttttagt agcgtttacat atattataaa tttttgtaac atttaaatgt 3480
tcactttatgt ggcattttct gtttctcatt ggtaataaagc atttctaata taagcaagta 3540
aaaatacaaaa tttaattact tagaaatgat aactaagatc acaatgactg tgtacccttt 3600
aaaatatctt ttaatttgag gaattacatt ctatctaaat tgtgtgtctg tttttgaaag 3660
gtgtgtgtctt tcaagtaggt ggtttttatt taaagtactt atttttttga ggttaaaata 3720
ttcagaaact acttttatgc attgtagcag taatcattta tggaatgctt ttgacatggt 3780
ttaacagggg aaccaaact tttaggcctg ttttttttgg ttgtccttgt tgttatattt 3840
ttaatggaaa tctgttactc aggctggtag cattaaatta gctttttacc ctcaactcaa 3900
gtagcaaaata tcatggagaa cttcagtgtt tagcaggctg ttaatttccc atctatgaat 3960
agcactttat gatgtaggct aaactaagca aataatgcag cgttatgtgg atcattatgc 4020
aataaagctg ctaaagtcac atttgtggtt gatactcaag ccttgtacaa gcacatatca 4080
tgtagacatt tttaaatgtg aagctttaac aggaatagga tagctgtttc aaaccttggtg 4140
caattgtaaa taaattggtt agctgtgttt ctttcatttt taataaaaaa ttaaaaaaaa 4200
aaaaaa 4206

```

&lt;210&gt; 107

&lt;211&gt; 1971

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7507432CB1

&lt;400&gt; 107

```

ggcgcgagg gcgcgctga gccgccatgt cggaggcgcg tggagagcca ggggccgggc 60
ctgaggctgg cgcccgatgc ttctgcactg cgggtcgcg cctggagccg ttcgtaatgc 120
gagagggtgc ggcgcggtgc gcggccacgc aggttgaata tatttcagga aagggttttt 180
tcaccacctg ttctgatttg aatatgttga agaaattaaa atctgcagaa agattatttt 240
tgctgattaa aaagcagttt ccacttatta tttcttctgt aagtaaagga aaaatattta 300
atgaaatgca aagacttata aatgaagatc caggaagtgt gttgaatgcc atttcaattt 360
ggaaaaatct tcttgaactt gatgcaaaaa aggaaaaact ttctcagaga gatgataacc 420
aactaaaaag aaaagtggga gaaaatgaaa tcattgcaaa gaaattaaaa atagaacaaa 480
tgcaaaagat agaagagaat agggactgcc agctggaaaa acaataaaaa gaagaaactc 540
tggagcaaag agattttacc actaaaagcg aaaagtttca agaagaagaa tttcagaatg 600
acatagagaa agcaattgat actcataatc agaattgactt gactttcaga gtatcttgtc 660
gctgcagtgg aactattgga aaggccttca ctgcacagga ggtaggaaaa gtaattggaa 720
ttgctattat gaaacacttt ggatggaaag cagacttgag gaatccacaa ttagagatct 780
ttatacatct aaatgacatt tactctgttg tggggattcc tgtgttcagg gtttccttag 840
ccagcagagc ttacatcaag acagctggac tgcgatctac aatagcgtgg gcaatggcat 900
ctctggctga catlaaggct ggtgcatttg ttttagatcc aatgtgtgga cttggaacaa 960
tacttttgga agctgctaaa gaatggccag atgtgtatta tgtaggtgct gatgtcagcg 1020
actcacagtt actaggtact tgggacaatc tgaaagctgc aggccttgag gataaaattg 1080
aattacttaa aatctctgtt atagaattgc cattgccttc agaaagtgtt gatattatta 1140
tttctgacat tccatttggg aaaaagttaa agttaggaaa agacatcaaa agcattctac 1200
aagaaatgga aagagtgtct catgttggcg gaaccattgt attgttgctt agtgaagatc 1260
accacaggcg cttacagat tgtaaagaga gcaacatccc tttcaattcc aaggacagtc 1320
acacagatga acctggaatt aaaaagtgtc tgaatcctga agaaaaaact ggtgcattca 1380
agacagcgtc aacttcattc gaagccagta accacaaatt cttagacaga atgtcaccat 1440
ttggctcctt ggtaccagtg gaatgctaca aagttagcct tggaaagaca gatgcgttca 1500

```

```

tatgtaaata taagaagtcg cactcttctg gactgtagca ggcttgctgc catcagccgg 1560
gttcaggccc ttgtaagtta gctgtacagc agaagtttgc tgtctctagg attcatatag 1620
caccacacaga gctccaagta accaaattcc cccaaagaca ggagggtgtgg ctgatgggta 1680
cacatgcttg gtaagagttt ctgctttata ttctaaagta cttaatgttt ttgagaaaaa 1740
agttaccgaa tatttttttag acttaagttt tatcttctgt cagaatgtga aaagttctgc 1800
ccagtttctg agcaaaaact tcaatttcaa aataaagggt ttcaggggtt cacttaagtg 1860
tatcacaatc aacaccagac gggatatttt aactatcaga ttttcgaaa tgtaaaaacc 1920
tttatgtgta acttcttttg atattaaact tgatttttga gcctgtaaaa a 1971

```

&lt;210&gt; 108

&lt;211&gt; 1365

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 5639578CB1

&lt;400&gt; 108

```

acgttgacgc tacctacctc tggcctggcc tctaccact gctgccctgg ccgctaccaa 60
catgagcctc cgggagagtc ctacacagccc cgctactctc gactatgctc tggaagaccc 120
gcaccagggc cagaggtccc gagaaaagag caaggcgaca gaggtaatgg cagatatgtt 180
tgatggccgc ttggagccaa tcgtgttccc accgccccgg cttccagagg aggggggtcg 240
gccccaggat cccgcagatg gtggccatac tttccacatc ttggtcgatg cgggtcgcag 300
tcatggagcc atcaaagcgg ggaggaagt gactccacca cccgcggagg gcctagaagc 360
ggcctctgcc tcgctgacaa ctgacggcag cctcaaaaat ggctttccgg gtgaagagac 420
gcacggccta ggtggggaga aggtctctaga aacctgtggg gcagggaggt cggagtctga 480
agtgattgca gagggaagg ccgaggacgt gaagcctgag gagtgtgcca tgttctcagc 540
cccagtggat gagaagccag gaggcgaaga aatggacgtg gcggaagaaa acagagcaat 600
agatgaggtg aacagggagg cagggcctgg gcccgggccc gggccccctg acgtgggtct 660
ccacctgaac cccctggagt ccatccagct ggaactggac tccgtgaatg cagaggctga 720
cagggcgctc ctgcaggtgg agcgcaggtt tgggcagatt catgaatact acctggagca 780
gaggaatgac atcattcgca atatccggg cttctgggtc actgctttcc gccaccaccc 840
acagctgtct gccatgatta gaggccaaga tgccgagatg ttaagctact taaccaattt 900
ggaagtgaag gagctcagac accctaggac aggttgcaag ttttaagttct tctttcaaag 960
aaacccttac ttcagaaaca agctgattgt aaagggtgat gaggtcagat ccttcggcca 1020
agtgggtgtc ttttccactc taatcatgtg gcgcgggggc catggacccc agtccttcat 1080
tcataggaac cgacatgtca tctgcagctt cttcacctgg ttttcagacc acagccttcc 1140
agagtccgac aggattgtct agattattaa agaggacctc tgggtcaaac cactgcagta 1200
ctacctgttg ggtgaagacg cccatagagc tagacgtcgc ctggtaaggg agccagtgga 1260
gatccccagg ccctttgggt tccagtgtgg ttaacctttg cctttggaat tactcctgca 1320
gaaggttgcc tgccaccacc agctggacct gagcttgggc aacag 1365

```

&lt;210&gt; 109

&lt;211&gt; 5261

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7509080CB1

&lt;400&gt; 109

```

ggagggtcaag atgggtggctg taaaactagg atccctgacg attgcttagc attaaggccc 60
gacatggaac cgggggtgtga cgagttcctg ccgccaccgg agtgcccggt ttttgagcct 120
agctgggctg aattccaaga cccgcttggc tacattgcga aaataaggcc catagcagag 180

```

```

aagtctggca tctgcaaaat ccgcccaccc gcggtattggc agcctccttt tgcagtagaa 240
gttgacaatt tcagattttac tcctcgcgtc caaaggctaa atgaactgga ggcccaaact 300
agagtgaat tgaactatctt ggatcagatt gcaaaattctt gggaaattca aggctcctct 360
ttaagattc ccaatgtgga gcggaagatc ttggacctct acagccttag taagcaatgt 420
aacacacacc cgtttgacaa tgaggtaaaa gataaggaat acaagcccca cagcatcccc 480
cttagacagt ctgtgcagcc ttcaaagttc agcagctaca gtcgacgggc aaaaaggcta 540
cagcctgatc cagagcctac agaggaggac attgagaagc atccagagct aaagaagtta 600
cagatatatg ggccagggtcc caaatgatg ggcttgggcc ttatggctaa ggataaggat 660
aagactgtgc ataagaaagt cacatgcccc ccaactgtta cggatgaagga tgagcaaagt 720
ggagggtggga acgtgtcatc aacattgtct aagcagcact tgagcctaga gccctgcact 780
aagacaacca tgcaacttcg aaagaatcac agcagtgtccc agtttattga ctcatatatt 840
tgccaagtat gctcccggtg ggatgaagat gataagcttc ttttctgtga tggctgtgat 900
gacaattacc acatctctct ctgtttacca ccccttcctg aaatccccag aggcattctg 960
agggtgcccc aatgtatctt ggcgaggtgt aaacagcctc ctgaagcttt tggatttgaa 1020
caggctaccc aggagtacag tttgcagagt tttgggtgaaa tggctgattc cttcaagtcc 1080
gactacttca acatgcctgt acatatggtg cctacagaac ttgtagagaa ggaattcttg 1140
aggctggtga gcagcattga ggaagacgtg acagttgaat atggagctga tattcattcc 1200
aaagaatttg gcagtggctt tcctgtcagc aatagcaaac aaaacttacc tcctgaggag 1260
aaggagtatg cgacagtgg ttggaacctg aatgtgatgc cagtgtctaga tcagtctgtt 1320
ctctgtcaca tcaatgcaga catctcaggc atgaaggtgc cctggctgta cgtgggcatg 1380
gttttctcag cattttgttg gcataattgag gatcactgga gttactctat taactatctg 1440
cattgggggtg agccgaagac ctggtatggt gtaccctccc tggcagcaga gcatttgagg 1500
gaggtgatga agatgtctgac acctgagctg tttgatagcc agcctgatct cctacaccag 1560
cttgtcactc tcatgaatcc caacactttg atgtcccatg gtgtgccagt tgtccgcaca 1620
aaccagtgtg caggggagtt tgtcatcact tttcctcgtg cttaccacag tggttttaac 1680
caaggctaca attttgctga agctgtcaac ttttgtactg ctgactggct acctgctgga 1740
cgccagtga ttgaactacta ccgcccgtc cgcgctatt gtgtcttctc ccacgaggag 1800
ctcatctgca agatggctgc cttcccagag acgttggtac tcaatctagc agtagctgtg 1860
cacaaggaga tgttcattat ggttcaggag gagcgacgtc tacgaaaggc ccttttgagg 1920
aaggcgctga cggagggtga gcgagaggct tttgagctgc tcccagatga tgaacggcag 1980
tgcatcaagt gcaagaccac gtgcttcttg tcagccctgg cctgtctacga ctgcccagat 2040
ggccttgtat gcccttccca catcaatgac ctctgcaagt gctctagtag ccgacagtac 2100
ctccggtatc ggtacacctt ggatgagctc cccaccatgc tgcataaact gaagattcgg 2160
gctgagctct ttgacacctg ggccaacaaa gtgcgagtgg ccttgagggt ggaggatggc 2220
cgtaaacgca gctttgaaga gctaagggca ctggagtctg aggctcgtga gaggagggtt 2280
cctaataagt agctgcttca gcgactgaag aactgcctga gtgaggtgga ggcttgatt 2340
gtcaagtcc tggggctggt cagtgggtcag gtggccagga tggacactcc acagctgacc 2400
ttgactgaac tccgggtcct tcttgagcag atgggcagcc tgccctgtgc catgcatcag 2460
attggggatg tcaaggatgt cctggaacag gtggaggcct atcaagctga ggctcgtgag 2520
gctctggcca cactgcctc tagtccagg ctattgcgtt cctgttgga gagggggag 2580
cagctgggtg tagagggtcc tgaagcccat cagcttcagc agcaggtgga gcaggcgcaa 2640
tggctagatg aagtgaagca ggccctggcc ccttctgtct acaggggtct tctgggtcatc 2700
atgcaggggc ttttggttat ggggtgccaag atagcctcca gcccttctgt ggacaaggcc 2760
cgggctgagc tgcaagaact actgaccatt gcagagcgct ggggaagaaa ggctcatttc 2820
tgcttgagg ccaggcagaa gcatccacca gccacattgg aagccataat tcgtgagaca 2880
gaaaacatcc ctgttcacct gcctaacatc caggctctca aagaagctct gactaaggca 2940
caagcttgga ttgctgatgt ggatgagatc caaatgggtg accactaccc ctgtctagat 3000
gacttgagg gcctggtggc tgtgggcccg gacctgcctg tggggctgga agagctgaga 3060
cagctagagc tgcaggattt gacagcacat tcctggagag agaaggcctc caagaccttt 3120
ctcaagaaga attcttgcta cacactgctt gaggtgaggt ctgagacctt gaccacagc 3180
ctcttcttca cctggcctgg ctgctgtgag atggcgcata taatgagaac atagattttt 3240
ttagtgggca cctgggtagg aaggagagg ttgtagttgt gagggagacc tggtcatttc 3300
ctgtatgtct gcctgcctgc ctgaggtgct ttgcccgtgt gcagacgctg gctcagacag 3360
caccaagcgt agccggtgga tggagaaggc gctgggggtt taccagtgtg acacagagct 3420
gctggggctg tctgcacagg acctcagaga cccaggctct gtgattgtgg ccttcaagga 3480
aggggaacag aaggagaagg agggatcct gcagctgcgt cgcaccaact cagccaagcc 3540

```

```

cagtccactg gcaccatccc tcatggcctc ttctccgact tctatctgtg tgtgtgggca 3600
gggtgccagct ggggtgggag ttctgcagtg tgacctgtgt caggactggg tccatgggca 3660
gtgtgtgtca gtgccccatc tcctcacctc tccaaagccc agtctcactt catctccact 3720
gctagccttg tgggaatggg acacaaaatt cctgtgtcca ctgtgtatgc gctcacgacg 3780
gccacgccta gagacaatcc tagccttgct ggttgccctg cagaggctgc ccgtgaggct 3840
gcctgagggt gaggcccttc agtgtctcac agagaggggc attggctggc aagaccgtgc 3900
cagaaaggct ctggcctctg aagatgtgac tgctctgttg cgacagctgg ctgagcttcg 3960
ccaacagcta caggccaaac ccagaccaga ggaggccctca gtctacactt cagccactgc 4020
ctgtgaccct atcagagaag gcagtggcaa caatatctct aagggtccaa ggctgctgga 4080
gaatggagac agtgtgacca gtctgagaa catggctcca ggaaagggtc ctgacctgga 4140
gctactgtcc tactgttgcc cgagttgac tggccctgtg ttggagctgc ctgaggcaat 4200
ccgggctccc ctggaggagc tcatgatgga aggggacctg cttgagggtga ccctggatga 4260
gaaccacagc atctggcagc tgctgcaggc tggacagcct ccagacctgg acagaattcg 4320
cacacttctg gagctggaag aatttgaaca tcaaggaggt cggacaagga gccgggctct 4380
ggagaggcga cggcggcggc agaagggtgga tcagggtaga aacgttgaga atcttggtca 4440
acaggagctt cagtcaaaaa gggctcggag ctcagggtat atgtctcagg tgggcccaga 4500
agaagaacat tatcaggaga aagcagaccg tgaaaatatg ttcttgacac cttccacaga 4560
ccacagccct ttcttgaaag gaaacaaaa tagcttacaa cacaaggatt caggctcttc 4620
agctgcttgt ccttctttaa tgcctttgct acaactctcc tactctgatg agcaacagtt 4680
gtgacagtgg caccaaaggc catttgtggt tgtttttgtt tgtttgttcc ttaaatccta 4740
ctatctcctg gcctggacct cagaaggagc tttttgccta tctataatct ttactgcca 4800
atttttgata tcctctctcc tagagttact gttaaaaggc tggttcgtaa agtccacacc 4860
ccgatgctca gaagtgtctt gccagcaaca ttctgctag catacaggag tgatttccta 4920
aaccagtttc attctagtct gaatagggac aaacaaatct tgaggagacc caagtgcgta 4980
cctttatctt tgccccacc accctctttc tgtacttcaa tttttgtttg tttttgttt 5040
ttttgtccct gtcataaaat attttggtgc ttcaaaactt gtaccttcat tgtacatcct 5100
tttcttttct ccccttgggt cttattataa aagaagacaa tgtacgttgt aattaccaa 5160
aagaataggg aaaaacaaga atttcatgac tctacctgtg gtctatcttt aatttcattt 5220
cttttgtaa aaataaaaca atgagtatgt ttggatacta a 5261

```

&lt;210&gt; 110

&lt;211&gt; 814

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7505899CB1

&lt;400&gt; 110

```

gctagtctta gatcgcgagg tccgcccttt tttttttttt tcctccctgt cgccaccgag 60
gtcgcacgag tgagacttct ccgccgcctc cgccgcagac gccgccgaga tgcgtactgt 120
cgctccctac ctgctggctg ccctaggggg caactcctcc ccagcgcca aggacatcaa 180
gaagatcttg gacagcgta tcaagtgtgt gaatggaaaa aacattgaag acgtcattgc 240
ccagggtatt ggcaagcttg ccagtgtacc tgctgggtgg gctgtagccg tctctgctgc 300
cccaggctct gcagcccttg ctgctgggtc tgccctgct gcagcagagg agaagaaaga 360
tgagaagaag gaggagtctg aagagtcaga tgatgacatg ggatttggcc tttttgatta 420
aattcctgct ccctgcaaa taaagccttt ttacacatct aaaaaaaaaa aaaaaagcg 480
gccgcttttt tttttttttt gagatgtgta aaaaggcttt atttgcaggg gagcaggaat 540
ttaatcaaaa aggccaaatc ccatgtcatc atctgactct tcagactcct ccttcttctc 600
atctttcttc tcctctgctg cagcaggggc agaaccagca gcaggggctg cagagcctgg 660
ggcagcagag acggctacag cccaccagc aggtacactg gcaagcttgc caataccctg 720
ggcaatgacg tcttcaatgt tttttccatt cagctcactg ataacgcggc ggcgtctgag 780
cgggagaagt ctcacgcgtg cgacctcggg ggcg 814

```

&lt;210&gt; 111

&lt;211&gt; 2637

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7505904CB1

&lt;400&gt; 111

```
aggctctcta cgggtgaggg tttgcgggga agatggagta tcccgcgccg gccacggtgc 60
aggccgcgga cggcggagcg gccgggcctt acagcagctc ggagttgctg gagggccagg 120
agccggacgg ggtgcgcttt gaccgcgaga gggcgccgcc cctgtgggaa gccgtgtccg 180
gtgcccagcc ggtgggtaga gaggaagtgg agcacatgat ccagaagaac caatgtctct 240
tcaccaaacac ccagtgtgtaag gtttgctgctg ccttgcttat ttctgagtc ccagaagctgg 300
cacattacca gagcaaaaaa catgccaaaca aagtgaagag atacctagca atccatggaa 360
tggagacatt aaagggggaa acgaagaagc tagactcaga tcagaagagc agcagaagca 420
aagacaagaa ccagtgtctgc cccatctgta acatgacctt ttcctcccct gtcgtggccc 480
agtcgcacta cctggggaag acccacgcaa agaacttaa gctgaagcag cagtccacta 540
aggtggaagc tggaaagggc taccctcgca aaacatgtaa gatagtgtg aactccatag 600
aacagtacca agctcatgtc agcggcttca aacacaagaa ccagtcacca aaaacagtgg 660
catcatccct gggccagatt ccaatgcaaa ggcaacccat tcagaaagac tcaaccacct 720
tggaaagacta gaggtgatcc tgcccagcat cccatattgg gccagccatg agccagcttc 780
ccgtgactgc tcagcccttg gctccctctt gctcgttgtt ctcaccagga aagtacacgg 840
gcctgaggca ggattgggccc acagacagcc tctcattggt ccgggctaata tcaactcctgc 900
tgctccccct tggcaggggt cctgtaggtc atgacagggg aggcaagggt attgagagac 960
tcgggggtctc gcgggggtgg agtttgaggg gtggctttcc ccatttccca acccctctgg 1020
gccttaggtg ctgaggcccc tgccacctgt ctttctctta aaggctcagtt ttggggccagt 1080
tcttgcaact aaagagcaga gatctctctg ggccttagac atttccagca aaacctggaa 1140
ctttcatgcc aaacctgggg cagggcagga aacagaggaa atggctgcaa catgggagct 1200
tggagctaact acgacactct gccttcccc agaagggtgca ggctttcctg agtcttagac 1260
cagatatggc cagttgcgca ggtttctgccc aactgtgaag tatcctcctg gagcagtgac 1320
acaatcttgg cggagcattg ctacccccgc tgccccctcc acagttcctg aatgggtgcta 1380
aggatctgca gcagttggca acgagctggg gctgggggcg gcctccatgt ccactgagat 1440
cataggacac tccaatgggg atgggacctt tccccctctc catcagaggt gctctgacct 1500
taggttacac gggaaagtgc cccacatgca agtctccctg aggggtctgc ccctaaaggc 1560
agactgcctc atgcccgtca gctgtgaggt tcattgtctac cctcggccct actagccctc 1620
tcttccccct tgtgcagcgg accacttgcc cagtttgctg tgggtgctagc cttccccatc 1680
atccaccggg tgattttctg gtcccaggga aagaaagaga gagctgatgc aggtttctac 1740
agtgaggaac aggcgtttcc caggccccac acccagattt ctctatcttt gctgtgtttt 1800
atggcctggg actgagtcca caggataga ttttctcttg taaccttgag acgagaattc 1860
caaggagtgt caccatcaga ggcttctctt cattgtgtca aagaagcccc tagctgctct 1920
cgtggcctcc tccccccact ccttatccct tcacctgtga aatgcctttg ctttgcatat 1980
tgtgtgtgga tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg tgtatgtgtg tgcttctgtg 2040
tgtgcctaata gctctgtctc ttggctactg aagcatccaa ataaagaatt tccctcatgg 2100
gccagactac aacttaacag acctaggggt ttgagccata ttgccattgg atctagtatg 2160
tctgtcttga cccctcacc ccacccatt cagacctcag tagtccccag tagtcacttt 2220
aggagtggg acagagtcca gcacctgtg aaagtctccc ttgagttcag gtgccctgaa 2280
ctcagtacag cctcatcttc tcatccacca cttctcctg cttctcctgt cacatagggc 2340
ctcatgaagg tctttaaagg gtttaagctg gggatctgct ttatgtgtga gagaactgtt 2400
ctggctgtta tgtaaaggag tgcaggagg ggcaggaaaag gagacaagga ggccagtatg 2460
gacactgtgt ccagtatggc ttccaggct aatcacttag ggcagagacc cagccttatt 2520
catctctact tctgatgtc catagtgggt tgtcctgaat tcttattgaa tgaccaatga 2580
cgtgtctcag gagaaatggg tctcctgccc agtctaaccg gatagtttgt ccggggc 2637
```

&lt;210&gt; 112

&lt;211&gt; 2699

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7509224CB1

&lt;400&gt; 112

```
ccgctttccc ggctccgctc ctgacgcgtc gtagacgttg gggagcggga aggcaacggc 60
agcgggatcg ggatgaacag cggcggcggc ttcggtttgg gcttaggctt cggcctcacc 120
cccacgtcgg tgattcaggt gacgaatctg tcgtcggcgg tgaccagcga gcagatgcgg 180
acgctttttt ctttcttagg agaaatcgag gagctgcggc tctaccccc ggacaacgca 240
cctcttgctt tttctccaa agtatgttat gttaagtctc gtgatccatc aagtgttggc 300
gtggcccagc atctaactaa cacgggtttt attgacagag ctctgatagt tgttccttgt 360
gcagaagtga tgcagccaac acgaacgggt gtcattgtga acacaacttt cgatcacgcg 420
gaaaacaccg tcctgggaaa gcgtccatgc ttgatatcgt ttggttcatt aacattaagt 480
ttccagtaca ggtaaaatcc cagaggaatc caaagccctc tctttatttg ctctgtctcc 540
aaccatgaca agtctgatgc ctggtgcagg attgcttcca ataccgacct caaatccttt 600
gactactctt ggtgtttcac ttagcagttt gggagctata ccagcagcag cactagacct 660
caacattgca acacttgag agataccaca gccaccactt atgggaaacg tggatccttc 720
caaaatagat gaaattagga gaacgggttt tgttggaat ctgaattccc agacaacgac 780
agctgatcaa ctactgaat tttttaaaca agttggagaa gtgaagtttg tgcggatggc 840
aggtgatgag actcagccaa ctcggtttgc ttttgtggaa tttgcagacc aaaattctgt 900
accaagggcc cttgctttta atggagtatt gtttggagac agggcactga aaataaatca 960
ctccaacaat gcaatagtaa aacccctga gatgacacct caggctgcag ctaaggagt 1020
agaagaagta atgaagcgag tacgagaagc tcagtcattt atctcagcag ctattgaacc 1080
agagtctgga aagagcaatg aaagaaaagg cggtcgatct cgttcccata ctcgctcaa 1140
atccaggtct agtcaaaat cccattctag aaggaagaag tcacaatcaa aacacaggag 1200
tagatcccat aatagatcac gttcaagaca gaaagacaga cgtagatcta agagcccaga 1260
taaaaaacgc tctaaatcaa gggagagacg gaagtcaagg agtcgttcgc attcacggca 1320
caagagaaaa gacactcgag aaaagatcaa ggaaaaaggaa agagtgaag agaaagacag 1380
ggaaaaggag agagagaggg aaaaggaccg ggaaaaagac aaggaaaagg acagagagag 1440
agaccgggac aaggacggg aaaaggaccg ggaaaaagac aaggaaaagg acagagagag 1500
agaacgggaa aaagagcatg agaaggatcg agacaaagag aaggaaaagg aacaggacaa 1560
agaaaaggaa cgagaaaaag acagatccaa agagatagat gaaaaagaa agaaggataa 1620
aaaatccaga acaccacca ggagttacaa tgcattcgca agatctcgta gttccagcag 1680
ggaaaggcgt aggaggagga gcaggagttc ttccagatcg ccaagaacat caaaaacat 1740
aaaaaggaaa tcttctagat ctccgtcccc caggagcaga aataagaagg ataaaaagag 1800
agaaaaagaa agggaccaca tcagtgaag aagagagaga gaacgttcaa cgtctatgag 1860
aaagattctt aatgatagag atgggaagga gaagtggag aagaacagta cttcacttaa 1920
agagaaagag cacaataaag aaccagattc aagtgtgagc aaagaagtag atgacaagga 1980
tgaccaagag actgaggaaa acaaaatata gcacaatgg aattgtcagc tgaatgaaga 2040
aaactctctt accaaaacaa aagcagtata ggaccgacaa gtgtacctct gactcaatg 2100
ctggaatcaa atccaaagct tttaattctc tcaacaagat gtaaacagga aagaaatcta 2160
gttgagcatg aagataggat ctaacagctt ttccagttgt tagatgactt tgtggccatc 2220
ttgttattga gtaagaaaat aaagcatgga catcatgaaa ataacagatg ttacccaaac 2280
tcattcttcta aaatctgtgc atttccatgg tggctgacac acttgatcatg tggctgtgta 2340
gtgtttgcc aagaaccattg caaataaatt gaacatcaaa gatccaagtt tgtactatcc 2400
ctaaagactg gagataagca ttggaggctc ttttaaaaaa tgctagttag tgaatttcgt 2460
atcgtttcta cttcttcttc ttattttcca atatatacag tttgatggat gtcgcttgga 2520
aattggtgca atatatacac acccttggtt gttgccaaat atgtaagaaa gttttaacat 2580
ttacttccca gggacttgcg cactggcgct caattcccc tattgggtct tcttcgccc 2640
ctgtcgacgc aactctctc taaaaaagtt gccataaat ccaaaaaggc gtaaaattc 2699
```

&lt;210&gt; 113

&lt;211&gt; 1569



&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7505922CB1

&lt;220&gt;

&lt;221&gt; unsure

&lt;222&gt; (1) ... (1569)

&lt;223&gt; a, t, c, g, or other

&lt;400&gt; 113

```
gaggacgaag ccagttctct ttttctggtc tgactggcct ggaaattccc cgagcctgan 60
cccgccccan agaaatcccc anccagcgtt tatagggcgc cgcngcggcg ctgcagagcc 120
cacagcagtc cgtgccgcgc tcccgcgcgc cagcgcccca gcgaggaagc agcgcgcagc 180
ccgcgcccca gcgcacccgc agcagcgccc gcagctcgtc cgcgccatgt tccaggcggc 240
cgagcgcccc caggagtggg ccatggaggg cccccgcgac gggctgaaga aggagcggct 300
actgaagcag cgccacgaca gcggcctgga ctccatgaaa gacgaggagt acgagcagat 360
ggtcaaggag ctgcaggaga tccgcctcga gccgcaggag gtgccgcgcg gctcggagcc 420
ctggaagcag cagctcacgc aggacgggga ctcgttcctg cacttggcca tcatccatga 480
agaaaaggca ctgaccatgg aagtgatccg ccagggtgaag ggagacctgg ccttctctca 540
cttcagaaac aacctgcagc agactccact ccacttggct gtgatcacca accagccaga 600
aattgctgag gcacttctgg gagctggctg tgatcctgag ctccgagact ttcgagga 660
tcccccccta caccttgccg gtgagcaggg ctgcctggcc agcgtgggag tcctgactca 720
gtcctgcacc accccgcacc tccactccat cctgaaggct accaactaca atggccacac 780
gtgtctacac ttagcctcta tccatggcta cctgggcata gtggagcttt tgggtgctct 840
gggtgctgat gtcaatgctc agctcacctg gggccgcccc agcaccggga tacagcagca 900
gctggggcag ctgacactag aaaaccttca gatgctgcca gagagtgagg atgaggagag 960
ctatgacaca gagtcaagat tcacggagtt cacagaggac gagctgccct atgatgactg 1020
tgtgtttgga ggccagcgtc tgacgttatg agcgcaaagg ggctgaaaga acatggactt 1080
gtatatattgt acaaaaaaaaa agttttatct ttctaaaaaa agaaaaaaga agaaaaaatt 1140
taaagggtgt acttatatcc aactgcaca ctgcctggcc caaaacgtct tattgtggta 1200
ggatcagccc tcattttgtt gcttttgtga actttttgta ggggacgaga aagatcattg 1260
aaattctgag aaaacttctt ttaaacctca cctttgtggg gtttttggag aaggttatca 1320
aaaatttcat ggaaggacca cattttatat ttattgtgct tcgagtgact gaccccgatg 1380
gtatcctgtg acatgtaaca gccaggagtg ttaagcgttc agtgatgtgg ggtgaaaagt 1440
tactacctgt caaggtttgt gttaccctcc tgtaaattgg gtacataatg tattgttggt 1500
aattattttg gtacttttat gatgtatatt tattaacacg atttttacaa aaaaaaaaaa 1560
aaaaaaaaa 1569
```

&lt;210&gt; 114

&lt;211&gt; 2852

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7507695CB1

&lt;400&gt; 114

```
gaagtaccgg tcagccagcc tcctgctgca gagccagtga actcaggtcg ggcttctcag 60
ctgcgcacat aggcagtact catcttggcc ctgggaagaa actcaagaag aagcttttga 120
aacataaagc ttggatgggg tttgacctct gcagggcagc gccagctat aggagtcc 180
ctgctgagca gagaagatga ctgcagaatt gagagaagcc atggccctag ccccatgggg 240
cccagtgaag gtgaaaaagg aggaggaaga agaagaaaac ttcccaggtc aggcattccag 300
```

```

ccaacaagtg cactccgaga acatcaaagt ctgggccccca gtgcaggggtc ttcagacagg 360
ccttgatgga tcagaagagg aagaaaaggg tcagaacata tcctggggata tggcggtagt 420
cctgaaagca actcaggagg cacctgctgc ttcaaccctt ggcagctact cattaccagg 480
gactctggcc aagagtgaga tactggagac tcatgggacc atgaactttc taggtgctga 540
aaccaagaac ctacagttac tggttccaaa aactgagata tgtgaggaag ctgaaaaaacc 600
cctcatcata tcagaaagaa tccagaaagc tgatcctcaa ggacctgagt taggagaagc 660
ttgtgaaaaa ggaaacatgt taaagaggca gagaataaag agagaaaaga aagatttcag 720
acaagtgata gtgaatgact gtcacttacc tgaaagcttc aaagaagagg aaaaccagaa 780
atgtaagaaa tctggaggaa aatatagcct taattctggc gctgttaaaa atccaaaaac 840
ccagcttggg caaaagcctt ttacgtgtag cgtgtgtggg aaaggattta gtcagagtgc 900
aaacctcgtt gtgcatcagc gaatccacac tggagagaaa ccctttgaat gtcatgagtgc 960
tgggaaggcc ttcattcaga gtgcaaacct cgttgtgcat cagagaatcc aactggaca 1020
gaaaccttat gtttgctcaa aatgtgggaa agccttcact cagagttcaa atctgactgt 1080
acatcaaaaa atccactcct tagaaaaaac ttttaagtgc aatgaatgtg agaaagcctt 1140
tagttacagc tcacaacttg ctcggcacca gaaagtccac attacggaaa aatgctatga 1200
atgtaatgaa tgtgggaaaa catttactag gagctcaaac ctcatgtgcc accagaggat 1260
ccacactggg gagaagccct ttgcctgtaa cgactgtggc aaagccttta cccagagtgc 1320
aaatcttatt gtacatcagc gaagccatac tggtgagaag ccatatgagt gtaaagagtgc 1380
tgggaaagcc tttagtgtgt tttcacacct tattgtgcac cagagaattc acactgcaga 1440
gaaaccttac gactgcagcg aatgtgggaa agccttcagt cagctctctt gccttattgt 1500
ccaccagaga attcacagtg gagatcttcc ttacgtgtgt aatgaatgtg ggaaggcctt 1560
cacatgtagc tcatacctac ttattcatca gagaattcat aatggagaaa aaccttacac 1620
atgtaatgag tgtgggaagg ccttcagaca gaggtcgagc ctaccgtgc accagagaac 1680
ccacactggg gagaagccct atgaatgtga gaagtgtggt gcagctttca tttccaactc 1740
acacctcatg cgataccata gaacctcatc tgttgaataa caagtaagga agaggaagac 1800
ctccagcatt ggtcataaacc ttctgcctcc ctaatgagac acctctttgc tgttttcttc 1860
ctcctctata aaagtgaggg ctgtgtcctt aaaagtata gttttcagga atgcagcaga 1920
agacacaaga aaagcatttc agaggctaata ttaaaacaaa aagtaagcac ctaaaggcaa 1980
ggactttgtt ttaactgtac ctttaagcggc catgttctct gaagtggaaat gtggccttcc 2040
taggaatggg tcgtacaaaag ctaagtggta atgatgctat ttggggaaag gtcttttttg 2100
cttaattttg ttttttaaaa ctctgatgat tgcttgagca acaggcaggt tatctgcctg 2160
gttgaattct ggttgaaccg tgtattctaa tatttctggg taagtgggta ctgggtaagg 2220
aaaccacttg gggtagcagt tcaacaattc acttacgaat gtttataagc tttccatttc 2280
ctaggttaatt ttttaaaagc cagtcaaaac aaaaacttta ctgaaaatgg acagaaatag 2340
gaaatggact ttttccttac tgtctatacc tcctgaacct tggattgtga aagatctggg 2400
gacctctggg tctgttctga ccattcccta gtctccatgg ccaagcactc aaggattgat 2460
ggacaccaca caccagctat attcatttgc caagatcaac agctccttct ccaaacaact 2520
caagcccca attcccatcg cattccattt ggggtgagatg caactaacag ccccttctcg 2580
gatgaaggag tcaaagaata aagcttgctc agaggcatgc ccagttgtc tctgctgtca 2640
ttctgctcaa gtccagccag gggctctcagt ttgggtttca ttctcacatt ctggatgggtg 2700
tttagaagaa atagtggagc tcttctcatt ctgagaagac caaagaggga aagtgatggg 2760
agaggggagg tggaaataag attcctaagg tcttaatttg attttttgag aaattcagag 2820
gtattggaat atgaaaaact tttagaagga cg 2852

```